10:28:45 OCA PAD INITIATION - PROJECT HEADER INFORMATION 10/17/88

Active

Project #: E-21-T10 Cost share #: E-21-393 Rev #: 0
Center # : R6583-T10 Center shr #: F6583-T10 OCA file #: 128
Work type : RES
Contract#: F30602-88-D-0025-0010 Mod #: Document : TO
Prime #: Contract entity: GTRC

Subprojects ? : N
Main project #:

Project unit: EE Unit code: 02.010.118
Project director(s):
PARIS D T EE

Sponsor/division names: AIR FORCE / GRIFFISS AFB, NY
Sponsor/division codes: 104 / 023

Award period: 880930 to 881130 (performance) 881231 (reports)

Sponsor amount	New this change	Total to date
Contract value	19,701.00	19,701.00
Funded	19,701.00	19,701.00
Cost sharing amount		442.00

Does subcontracting plan apply ?: Y

Title: RADIATION FIELD OF OFFSET DUAL REFLECTOR ANTENNAS

PROJECT ADMINISTRATION DATA

OCA contact: Brian J. Lindberg 894-4820

Sponsor technical contact

ROBERT A. SHORE
DEPARTMENT OF THE AIR FORCE
ROME AIR DEVELOPMENT CENTER/EEAS
GRIFFISS AFB, NY 13441-5700

Sponsor issuing office

GERARD J. BROWN/PKRM
(315)330-7060
ROME AIR DEVELOPMENT CENTER
DIRECTORATE OF CONTRACTING (PKRM)
GRIFFISS AFB, NY 13441-5700

Security class (U,C,S,TS) : U
Defense priority rating : DO-A7
Equipment title vests with: Sponsor
NONE PROPOSED OR ANTICIPATED.

ONR resident rep. is ACO (Y/N): Y
GOVT supplemental sheet
GIT

Administrative comments -
DELIVERY ORDER FULLY FUNDS TASK NO. E-8-7161 (UNIV OF SOUTHERN CALIFORNIA).
$1,746 COST-SHARING REQUIRED BY SUBCONTRACTOR.

GEORGIA INSTITUTE OF TECHNOLOGY
OFFICE OF CONTRACT ADMINISTRATION

NOTICE OF PROJECT CLOSEOUT

Closeout Notice Date 03/30/90

Project No. E-21-T10 Center No. R6583-T10

Project Director JOY E B School/Lab EE

Sponsor AIR FORCE/GRIFFISS AFB, NY

Contract/Grant No. F30602-88-D-0025-0010 Contract Entity GTRC

Prime Contract No.

Title RADIATION FIELD OF OFFSET DUAL REFLECTOR ANTENNAS

Effective Completion Date 890630 (Performance) 890731 (Reports)

Closeout Actions Required:	Y/N	Date Submitted
Final Invoice or Copy of Final Invoice	Y	
Final Report of Inventions and/or Subcontracts	Y	
Government Property Inventory & Related Certificate	Y	
Classified Material Certificate	Y	
Release and Assignment	Y	
Other	N	

Comments

Subproject Under Main Project No.

Continues Project No.

Distribution Required:

Project Director	Y
Administrative Network Representative	Y
GTRI Accounting/Grants and Contracts	Y
Procurement/Supply Services	Y
Research Property Managment	Y
Research Security Services	Y
Reports Coordinator (OCA)	Y
GTRC	Y
Project File	Y
Other	N
	N

NOTE: Final Questionnaire sent to PDPI.

E-21-T10

GEORGIA INSTITUTE OF TECHNOLOGY
SCHOOL OF ELECTRICAL ENGINEERING
ATLANTA, GEORGIA 30332

TELEPHONE: (404) 894-

January 26, 1990

MEMORANDUM

TO: Jim Wasielewski. RADC/RBC

FROM: Lori Rose, Research Administrator

SUBJECT: FINAL REPORT - TASK E-8-7161
CONTRACT # f30602-88-D-0025

Attached is a draft copy of subject above. A typed copy will be submitted as soon as completed. Please call if you have questions or need additional information.

DEPARTMENT OF ELECTRICAL ENGINEERING
ELECTROPHYSICS

October 16, 1989

Ms. Kathy Knighton
Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

Dear Ms. Knighton:

At long last we have completed the final report of Award No. E-21-T10-81. It is currently in final typing. In the meantime I am forwarding you the rough copy in case you wish to make use of it. The final version will be mailed to you as soon as it is ready.

Sincerelv.

W.V.T. Rusch
Professor, Electrical Engineering

WVTR/drc
cc: A. Prata, Jr.

Efficient Computation of the Radiation Field of Classical Offset Dual-reflector Antennas Excited by Array Feeds

by

W. V. T. Rusch and A. Prata, Jr.

July 1989

This work was supported by the Rome Air Development Center through contract F30602-88-D-0025, E-8-7161

Department of Electrical Engineering
University of Southern California
Los Angeles, California 90007

Abstract

This report investigates the evaluation of the radiation characteristics of classical offset dual-reflector antennas excited by an array of feed elements. The subreflector is assumed illuminated by an array of identical feeds, with each element arbitrarily located, pointed, excited, and linearly polarized. The subreflector scattering is evaluated using either one of two methods. In the first the subreflector-induced currents are determined using the Physical-Optics (PO) approximation in conjunction with the array incident fields. Integration of these currents over the near-field free space dyadic Green's function then yields the subreflector scattered field. In the second the Geometrical Theory of Diffraction (GTD) is used to obtain the subreflector diffracted field due to each array element, and the total field is then obtained by summing up the contributions of each element. Once the subreflector scattering has been determined, physical-optics is used to obtain the main-reflector-induced currents, and by integrating these currents over the free-space dyadic Green's function the main reflector far-zone field is evaluated. Blocking effects of any kind are not taken into consideration. Vector fields are used throughout this work, and no approximations are made on angles and vectors.

The main objectives of this report are

to present the numerical tools required to analyze array-fed offset dual-reflector antennas, and to document a computer program implementation of these tools. The program is named SORADC, and this report contains detailed information on its operation, together with several sample runs to demonstrate its usage.

CONTENTS

Page

LIST OF ILLUSTRATIONS

6.3 – Cassegrain Antenna from the Literature Radiation Patterns. GTD Results

6.4 – Cassegrain Antenna from the Literature Radiation Patterns. PO Results

6.5 – Gregorian Antenna

6.6 – Gregorian Antenna Radiation Patterns. GTD Results

6.7 – Gregorian Antenna Radiation Patterns. PO Results

6.8 – Radiation Pattern of Array Fed Cassegrain Antenna. GTD Results

6.9 – Radiation Pattern of Array Fed Cassegrain Antenna. PO Results

6.10 – Radiation Pattern of Array Fed Gregorian Antenna. GTD Results

6.11 – Radiation Pattern of Array Fed Gregorian Antenna. PO Results

1- Introduction

The work described in this report was undertaken with the specific objective of developing an efficient computational procedure for calculating the radiation characteristics of classical offset dual-reflector antennas excited by an array of feed elements. By a classical configuration it is here understood that the antenna employs a paraboloidal main-reflector excited by a confocal subreflector, this last being either convex hyperboloidal or concave ellipsoidal. Figure 1.1 shows schematicaly the geometry considered, with a concave subreflector. The antenna radiates energy through a double scattering procedure where the main reflector illumination is provided by the subreflector scattering of the array feed.

In a previous report the problem of determining the feed array excitation coefficients was addressed, in order to have the dual-reflector antenna radiate a prescribed far-zone pattern [1.1]. In here the inverse problem is considered, namely the determination of the dual-reflector antenna radiation characteristics once the array excitation coefficients are given. The present work complements the one of ref. [1.1], since through it the antenna final radiation characteristics can be evaluated to confirm that indeed the prescribed performance was achened. This is a necessary test if one considers that, in ref. [1.1], the procedure for obtaining the array excitation coefficients involves sampling the

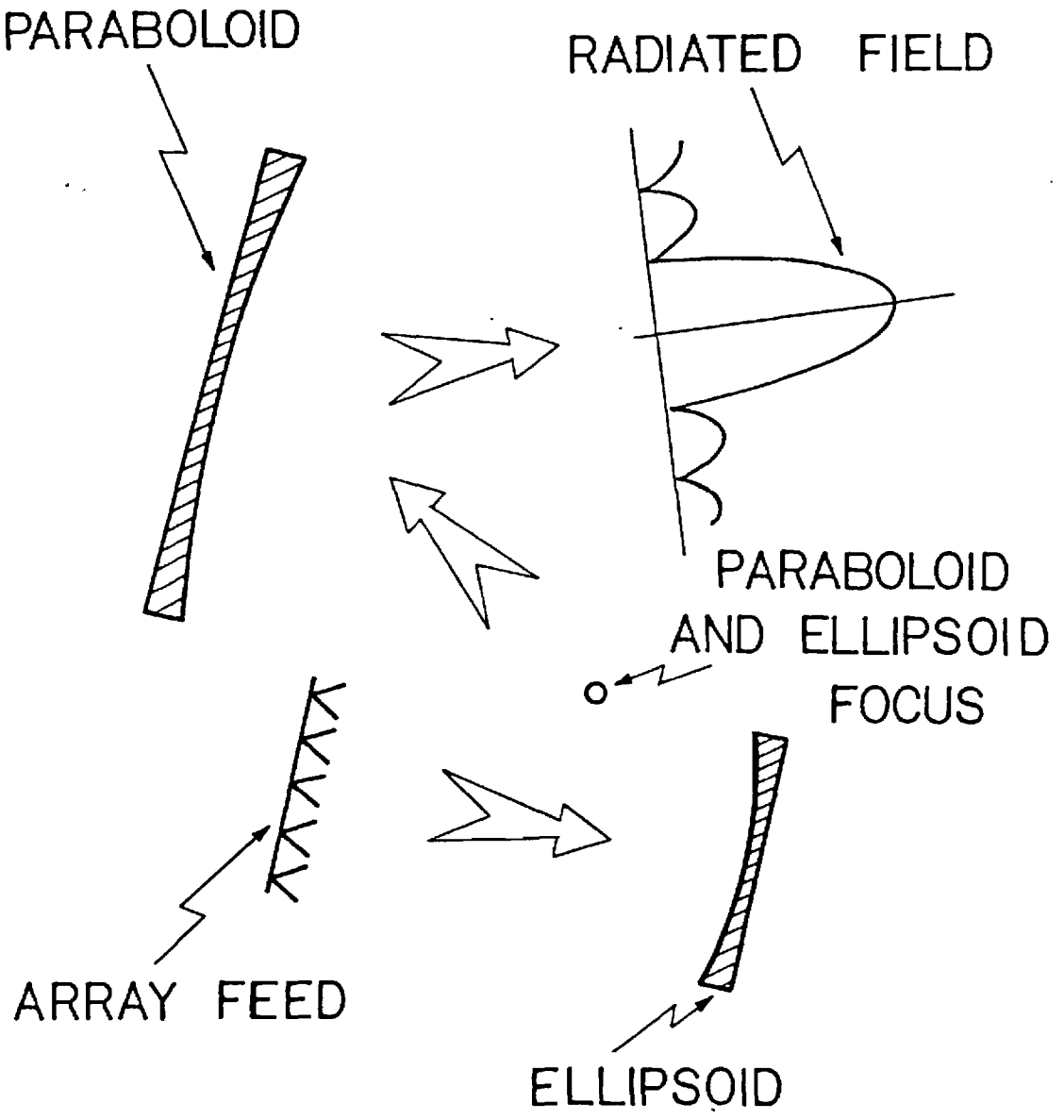


Fig. 1.1 - Classical dual-reflector antenna excited by an array feed (Gregorian geometry).

dual-reflector antenna focal-region fields at selected points, a technique that introduces several approximations and hence requires verification.

Our primary concern in this work is with antennas that generate a reasonably colimated radiation emanating from the main-reflector (i.e. high-gain antennas). Also, we are primarily interested in radiation pattern features ocurring in the vicinity of the radiated energy far-field peak direction. Under these conditions the dual-reflector antenna radiation characteristics is dominated by the main-reflector scattering, and hence the evaluation of only the paraboloidal reflector scattering yields an excellent approximation for the complete antenna radiation performance. We are then taking full advantage of this fact in this report, and hence we are evalvating the subreflector scattering only to determine the main-reflector illumination (the subreflector far-field scattering is being ignored).

In order to evaluate the main- and subreflector scattered fields, the procedure of integrating their corresponding induced surface current densities over the free-space dyadic Green's function has been adopted (the so-called PO method) [1.2]. This procedure is particularly well suited for the main-reflector scattering computation since, for our antenna geometry and pattern region of interest, the PO integrand is only mildly oscillatory, yielding then a numerically efficient algorithm. For the subreflector scattering evaluation however, the PO integrand is highly oscillatory,

and hence the evaluation of the corresponding surface integral can be very inefficient numerically. In other to provide an alternative to overcame this problem another method to evaluate the subreflector scattering has also been explored, one using geometrical theory of diffraction (the so called GTD method) [1.3]. Since GTD is based on the stationary phase principle, it is well suited to replace the PO method whenever a highly oscillatory integrand is encountered, which is our case.

At a first look it may appear that, between the GTD and PO methods, the second is always more efficient in evaluating the subreflector scattering. This is indeed true for a subreflector excited by a single feed element [1.4], but this is not our case since our antenna system employs an array feed with typically hundreds of elements. In orden to appreciate the significance of this point one should recall that the GTD method is based on ray-fields produced by reflection and diffraction points. In any GTD computation the reflection and diffraction points must then be searched for, and their corresponding ray-fields evaluated, and this procedure must be repeated for each new feed element observation point pair. Whenever an array feed is used this becomes a time consuming task and tends to offset the numerical advantage of GTD over PO. Which method is more efficient depends then on the specific antenna being evaluated - the GTD method being favored, over PO, by large subreflectors and/or

small number of feed array elements. Due to this fact both methods were considered in this work.

In the next pages the theory necessary to analyze classical dual-reflector antennas excited by array feeds is discussed and the final result of this work, the computer code SORADC, is described. SORADC is a general purpose computer program capable of evaluating the radiation characteristics of the aforementioned antenna systems — it summarizes the work here presented.

Considering the objective of the task described by this report, which is only to develop analytical and numerical tools, no direct atempt has been made to work on the design trade-offs of the antennas considered. This would be the next step and this work is intended to provide the necessary tools for it.

This report and corresponding computer code, is intended to be used either independently or in association with the work (and code) described in ref. [1.1]. In fact, the capabilities of code SORADC exceed the ones required by an usage associated only with the work of ref. [1.1]. Since it is felt that this is an useful characteristic for other potential applications we made all possible efforts to keep the present work general. As a consequence of this some data used to specify the antenna system is described slightly differently in codes SORADC and OFRADC. However, this presents no major problem to the user of both codes since an additional subroutine has been added to update code OFRADC to handle this

data difference. Apart from providing all the required data conversions, this subroutine also generates part of the data needed to run code SORADC when the focal region Field information (provided by code OFRADC is being used. In this way the full compatibility of both codes is preserved, and the capabilities of code SORADC are not compromised. This report (mainly through its Appendix I) provides also full information on this code OFRADC update.

This report has been divided in nine sections. Section 2 discusses the antenna geometry. Section 3 characterizes the feed array utilized to excite the system. Section 4 deals with the computation of the sub-reflector scattering. Section 5 presents the computation of the main-reflector scattering. Section 6 describes the computer program SORADC and also contains several sample runs of the code. And Section 9, the appendices, contains the above mentioned code OFRADC upgrade, as well as the FORTRAN source code for program SORADC.

Throughout this work the $e^{+j\omega t}$ time dependence has been assumed implicit for all fields and currents. All dimensions are assumed to be in wavelengths, unless otherwise explicitly stated. And a bar on top of a variable has been used when it is a vector, and a hat when it is an unit vector.

2- Antenna Geometry

The two classical dual-reflector geometries considered in this work are the offset Cassegrain configuration, shown in Fig. 2.1, and the Gregorian one of Fig. 2.2. Also shown in these figures is the principal coordinate system for this work, the X, Y, Z cartesian system. Both antenna geometries have symmetry about the Y=0 plane. In order to help visualize the reflector configurations, and to define a few other useful coordinate systems, in Fig 2.3 a perspective view of the Gregorian geometry is also shown.

In both reflector systems the main-reflector is an offset section of a paraboloid of revolution, described by the equation

$$z = f(x,y) = \frac{x^2+y^2}{4F} - F. \qquad (2.1)$$

Where F is the paraboloid focal length. The paraboloid axis is coincident with the z-axis, and its focus is located at the principal coordinate system origen. The projection of the main-reflector rim on the xy-plane is assumed to be a circle of diameter D, offset by a distance d_o from the z-axis.

An useful coordinate system for computing the main-reflector scattering is the localized R_A, ϕ_A, Z_A main-reflector cylindrical system shown in Fig. 2.3.

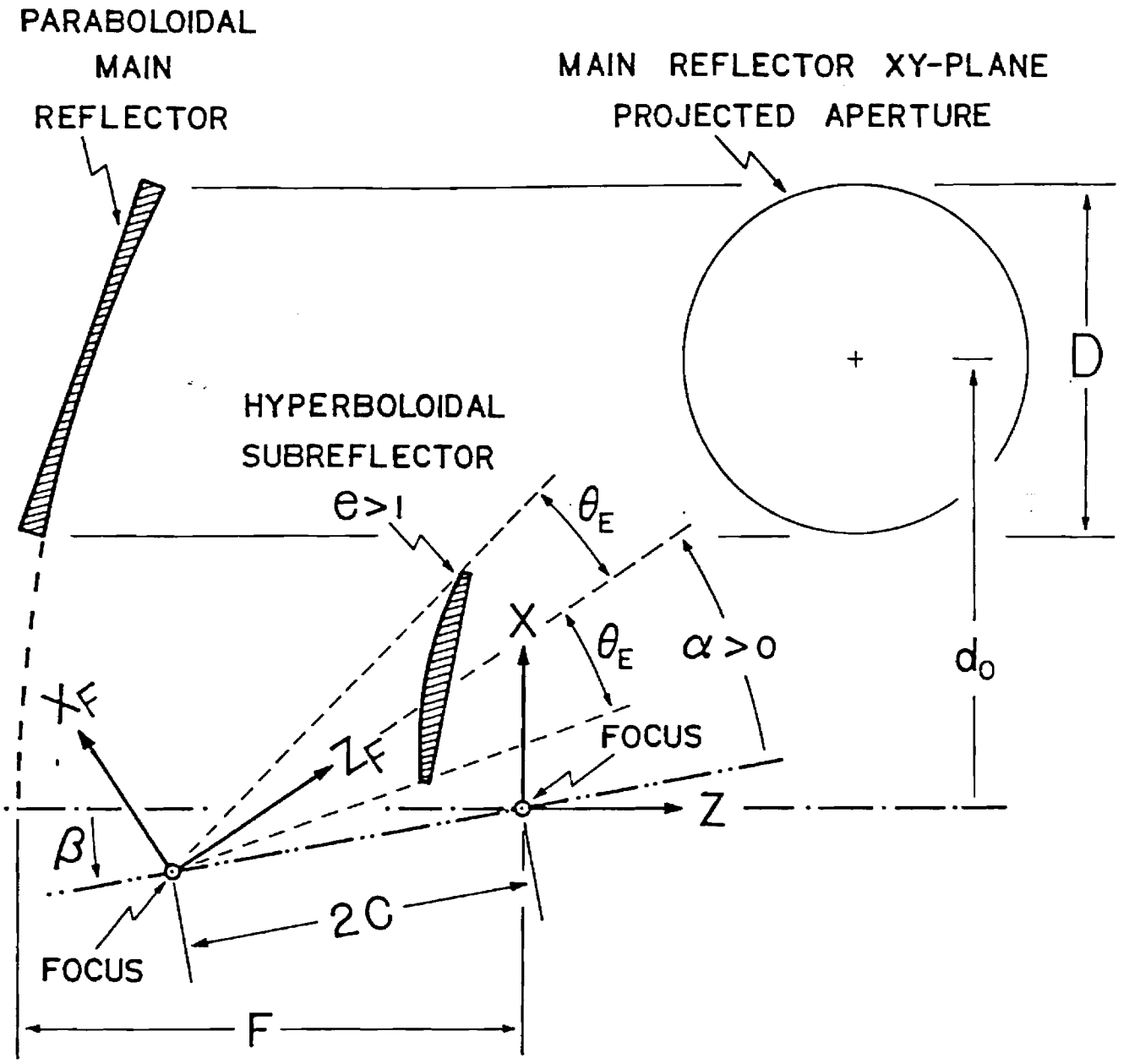


Fig. 2.1 - Classical Cassegrain geometry.

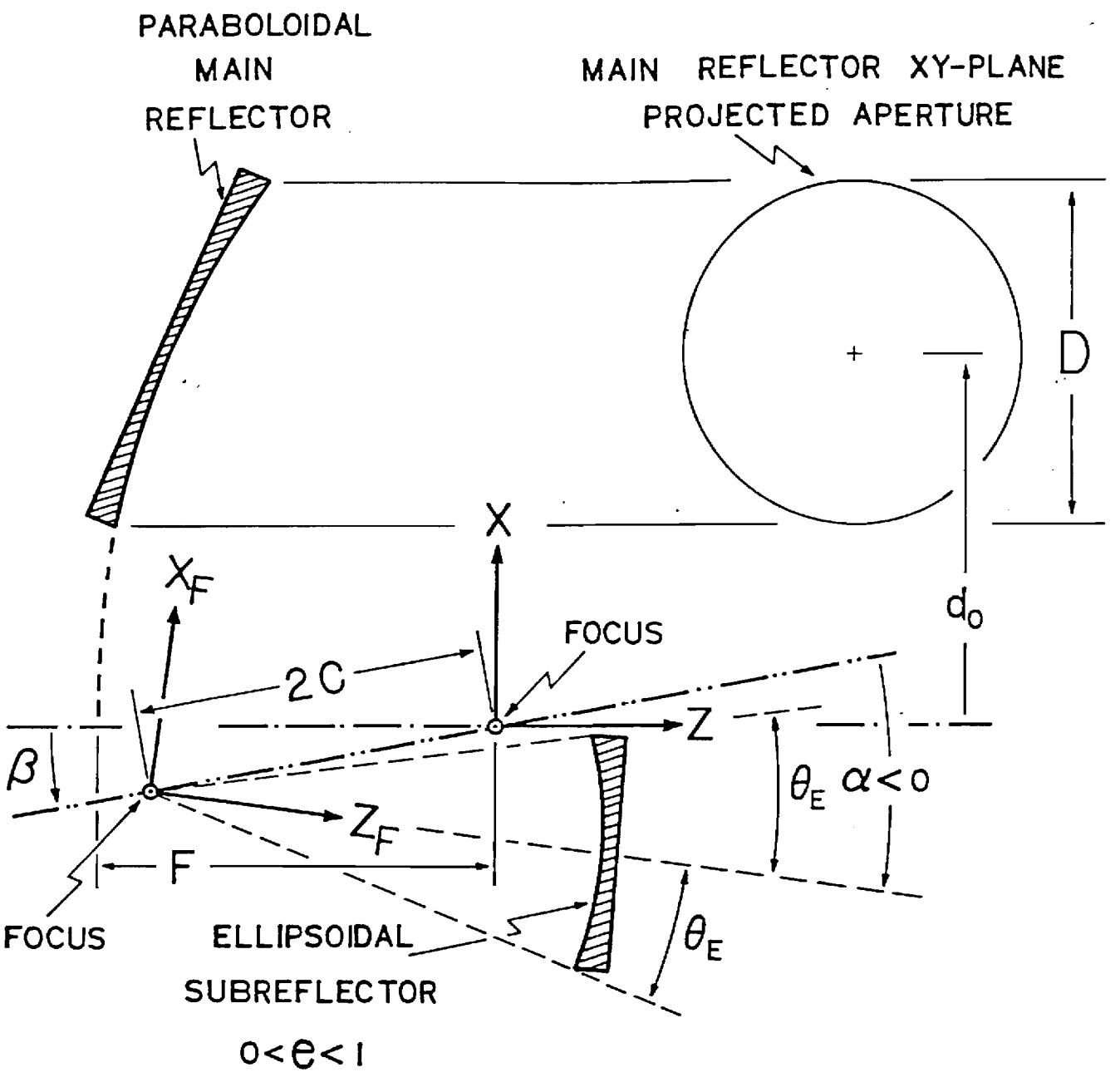


Fig. 2.2- Classical Gregorian geometry.

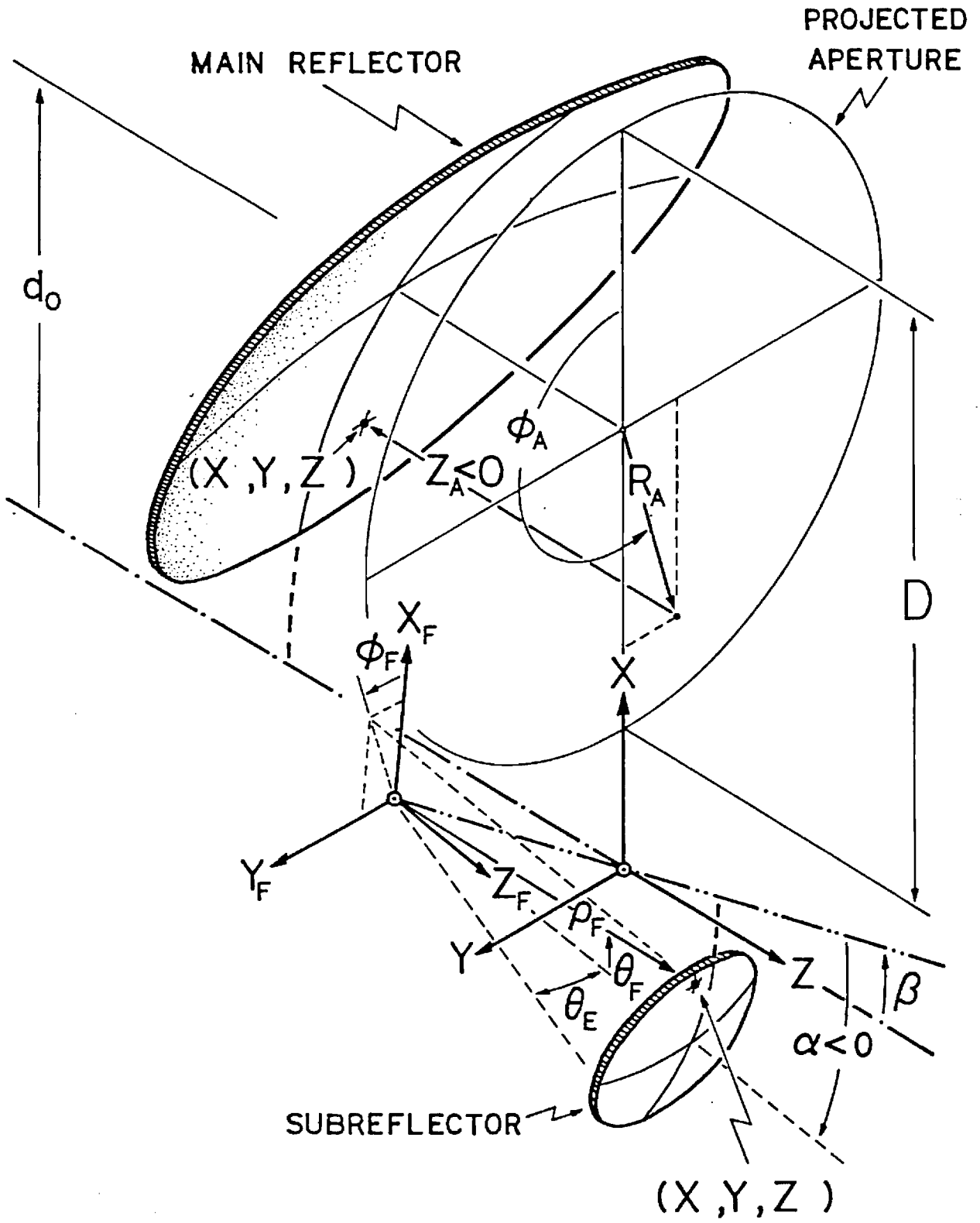


Fig. 2.3 - Classical Gregorian geometry (Perspective view).

It relates to the principal system through

$$X = R_A \cos\phi_A + d_o, \tag{2.2a}$$

$$Y = R_A \sin\phi_A, \tag{2.2b}$$

$$Z = Z_A, \tag{2.2c}$$

and will be used in a coming section of this report.

In order to define the subreflector surface, and evaluate its scattering, two other coordinate systems are also shown in Figs. 2.1-2.3. They are the X_F, Y_F, Z_F subreflector cartesian system and the ρ_F, θ_F, ϕ_F subreflector spherical system. The principal and subreflector cartesian systems relate to each other through

$$X = X_F \cos(\beta+\alpha) + Z_F \sin(\beta+\alpha) - 2c \sin\beta, \tag{2.3a}$$

$$Y = Y_F, \tag{2.3b}$$

$$Z = -X_F \sin(\beta+\alpha) + Z_F \cos(\beta+\alpha) - 2c\cos\beta, \tag{2.3c}$$

and

$$X_F = X \cos(\beta+\alpha) - Z \sin(\beta+\alpha) - 2c \sin\alpha, \tag{2.4a}$$

$$Y_F = Y, \tag{2.4b}$$

$$Z_F = X \sin(\beta+\alpha) + Z \cos(\beta+\alpha) + 2c\cos\alpha. \tag{2.4c}$$

And the subreflector cartesian and spherical systems relate to each other in the costumary way, namely

$$X_F = \rho_F \sin\theta_F \cos\phi_F, \tag{2.5a}$$

$$Y_F = \rho_F \sin\theta_F \sin\phi_F, \tag{2.5b}$$

$$Z_F = \rho_F \cos\theta_F. \tag{2.5c}$$

Another useful set of equations is the one relating the principal cartesian system and the subreflector spherical system. They can b obtained by substituting the above equations in Eqs. (2.3), and are given by

$$X = \rho_F \sin\theta_F \cos\phi_F \cos(\beta+\alpha) + \rho_F \cos\theta_F \sin(\beta+\alpha) - 2c\sin\beta, \tag{2.6a}$$

$$Y = \rho_F \sin\theta_F \sin\phi_F, \tag{2.6b}$$

$$Z = -\rho_F \sin\theta_F \cos\phi_F \sin(\beta+\alpha) + \rho_F \cos\theta_F \cos(\beta+\alpha) - 2c\cos\beta. \tag{2.6c}$$

Having presented these two subreflector coordinate systems, we are now in position to treat the subreflector surface specification. As shown in Figs. 2.1 and 2.2, the subreflector is a section of either a convex hyperboloid or a concave ellipsoid. In any case one of the subreflecto foci is coincident with the paraboloid focus, and the other is then the antenn system focus (located at the point with $X_F = Y_F = Z_F = 0$). The angle β allows then for the subreflector axis to be depresse

with respect to the paraboloid axis, an useful cross-polarization reducing feature [2.1]. And the angle α defines the $\theta_F = 0$ direction of the subreflector spherical system. The surface of both subreflectors are described by the equation

$$\rho_F = g(\theta_F, \phi_F) = \frac{\frac{c}{e}(e^2-1)}{e\cos\theta_s - 1}, \qquad (2.7)$$

where θ_s is the angle between the vector ρ_F and the subreflector axis, given in a convenient form by

$$\cos\theta_s = \cos\theta_F \cos\alpha - \sin\theta_F \cos\phi_F \sin\alpha. \qquad (2.8)$$

In Eq. (2.7) "e" is the subreflecton eccentricity and "2c" the interfocal distance — $e>1$ for a hyperboloidal subreflecton and $0<e<1$ for an ellipsoidal one. These two parameters fully characterize the subreflecton shape.

The actual subreflecton used in the antenna system is only a section of the one described by Eq. (2.7). As shown in Fig. 2.3, its rim is assumed to lie on the surface of a circular cone, and this cone has vertex located at the system focus, center direction specified by the angle α, and vertex semi-angle equal to θ_E.

At this point the main- and subreflecton of our antenna system are fully defined, and since the feed array is treated in the next section, the discussion

of the antenna geometry is complete. But, before closing this section, the equations of the main- and subreflector surface unit normals will be given - they are used in computing the corresponding scattered field using PO.

The main-reflector normal $\bar{N}_R$ is obtained from Eq. (2.1) by use of the gradient operator in cartesian coordinates then:

$$\bar{N}_R = \nabla \left[Z_R - f(X_R, Y_R) \right]. \tag{2.9}$$

And hence

$$\hat{N}_R = \frac{1}{\sqrt{\left(\frac{\partial f}{\partial X_R}\right)^2 + \left(\frac{\partial f}{\partial Y_R}\right)^2 + 1}} \left[-\hat{X}\frac{\partial f}{\partial X_R} - \hat{Y}\frac{\partial f}{\partial Y_R} + \hat{Z} \right], \tag{2.10}$$

where

$$\frac{\partial f}{\partial X_R} = \frac{X}{2F}, \tag{2.11a}$$

$$\frac{\partial f}{\partial Y_R} = \frac{Y}{2F}. \tag{2.11b}$$

In the above X_R, Y_R, and Z_R are the principal coordinates of the main-reflector surface point, and $\hat{N}_R$ its corresponding unit normal (pointing into the positive Z-space).

Similarly to the main-reflector, the subreflector normal $\bar{N}_S$ can be obtained from Eq. (2.7) using the gradient operator in spherical coordinates, then:

$$\bar{N}_S = \hat{\rho}_F - \hat{\theta}_F \frac{1}{\rho_F} \frac{\partial g}{\partial \theta_F} - \hat{\phi}_F \frac{1}{\rho_F \sin\theta_F} \frac{\partial g}{\partial \phi_F} . \qquad (2.12)$$

And hence

$$\hat{N}_S = \frac{-1}{\sqrt{1+\left(\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right)^2+\left(\frac{1}{\rho_F \sin\theta_F}\frac{\partial g}{\partial \phi_F}\right)^2}} \left[\hat{\rho}_F - \hat{\theta}_F \frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F} - \hat{\phi}_F \frac{1}{\rho_F \sin\theta_F}\frac{\partial g}{\partial \phi_F}\right], \qquad (2.13)$$

where

$$\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F} = \frac{e(\sin\theta_F \cos\alpha + \cos\theta_F \cos\phi_F \sin\alpha)}{e\cos\theta_s - 1}, \qquad (2.14a)$$

$$\frac{1}{\rho_F \sin\theta_F}\frac{\partial g}{\partial \phi_F} = \frac{-e\sin\alpha \sin\phi_F}{e\cos\theta_s - 1} . \qquad (2.14b)$$

In the above ρ_F, θ_F, and ϕ_F are the spherical coordinates of the subreflector surface point, and the unit normal $\hat{N}_S$ points into the $\rho_F = 0$ space.

The $\hat{N}_S$ given by Eq. (2.13) is in spherical coordinates, to transform it into the principal coordinate system recall that:

$$\hat{\rho}_F = \hat{X}_F \sin\theta_F \cos\phi_F + \hat{Y}_F \sin\theta_F \sin\phi_F + \hat{Z}_F \cos\theta_F, \qquad (2.15a)$$

$$\hat{\theta}_F = \hat{X}_F \cos\theta_F \cos\phi_F + \hat{Y}_F \cos\theta_F \sin\phi_F - \hat{Z}_F \sin\theta_F, \qquad (2.15b)$$

$$\hat{\phi}_F = -\hat{X}_F \sin\phi_F + \hat{Y}_F \cos\phi_F, \qquad (2.15c)$$

AND [FROM Eqs. (2.4)],

$$\hat{X}_F = \hat{X}\cos(\beta+\alpha) - \hat{Z}\sin(\beta+\alpha), \tag{2.16a}$$

$$\hat{Y}_F = \hat{Y}, \tag{2.16b}$$

$$\hat{Z}_F = \hat{X}\sin(\beta+\alpha) + \hat{Z}\cos(\beta+\alpha), \tag{2.16c}$$

which, upon substitution in Eq. (2.13) finally yields:

$$\hat{N}_S = \frac{-1}{\sqrt{1+\left(\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right)^2+\left(\frac{1}{\rho_F \sin\theta_F}\frac{\partial g}{\partial \phi_F}\right)^2}}\left[B_1\hat{X} + B_2\hat{Y} + B_3\hat{Z}\right] \tag{2.17}$$

where

$$B_1 = \cos(\beta+\alpha)\left[\cos\phi_F\left(\sin\theta_F - \cos\theta_F\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right) + \sin\phi_F\frac{1}{\rho_F\sin\theta_F}\frac{\partial g}{\partial \phi_F}\right] +$$

$$+\sin(\beta+\alpha)\left[\cos\theta_F + \sin\theta_F\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right] \tag{2.18a}$$

$$B_2 = \cos\phi_F\left[\sin\theta_F - \cos\theta_F\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right] - \cos\phi_F\frac{1}{\rho_F\sin\theta_F}\frac{\partial g}{\partial \phi_F} \tag{2.18b}$$

$$B_3 = \cos(\beta+\alpha)\left[\cos\theta_F + \sin\theta_F\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right] -$$

$$-\sin(\beta+\alpha)\left[\cos\phi_F\left(\sin\theta_F - \cos\theta_F\frac{1}{\rho_F}\frac{\partial g}{\partial \theta_F}\right) + \sin\phi_F\frac{1}{\rho_F\sin\theta_F}\frac{\partial g}{\partial \phi_F}\right] \tag{2.18c}$$

3 - Feed Array Characterization

The first step towards computing the dual-reflector antenna radiated field is the specification of the feeds used for its excitation. In this work the antenna excitation is assumed to be provided by an array of equal linearly polarized feeds, with each element arbitrarily located, pointed, and excited.

In Fig. 3.1 the parameters used to locate and orient each element of the array feed are shown. The feed element, represented as a conical horn, is positioned at the point X_E, Y_E, Z_E in the principal coordinate system. $\bar{R}_F$ is the vector from the feed to the field point P, and $\hat{R}'_F$ its respective unit vector. ϕ_{Po} is the feed polarization angle (the same for all elements); when $\phi_{Po}=0$ the electric field $\bar{E}$ is polarized in the $\hat{X}'_F$ direction, when $\phi_{Po}=90°$ in the $\hat{Y}'_F$ direction and so on (X'_F, Y'_F, Z'_F is the feed cartesian coordinate system, and R'_F, θ'_F, ϕ'_F its associated spherical system).

To allow each feed to be aimed at any arbitrary direction two other cartesian coordinate systems are also shown in Fig. 3.1: the x', Y', Z' and the X'', Y'', Z'' systems. The five coordinate systems shown in Fig. 3.1 are related by the following equations:

$$X' = X - X_E, \tag{3.1a}$$
$$Y' = Y - Y_E, \tag{3.1b}$$
$$Z' = Z - Z_E, \tag{3.1c}$$

$$X'' = X', \tag{3.2a}$$
$$Y'' = Y' \cos A_z + Z' \sin A_z, \tag{3.2b}$$
$$Z'' = Y' \cos A_z + Z' \cos A_z \tag{3.2c}$$

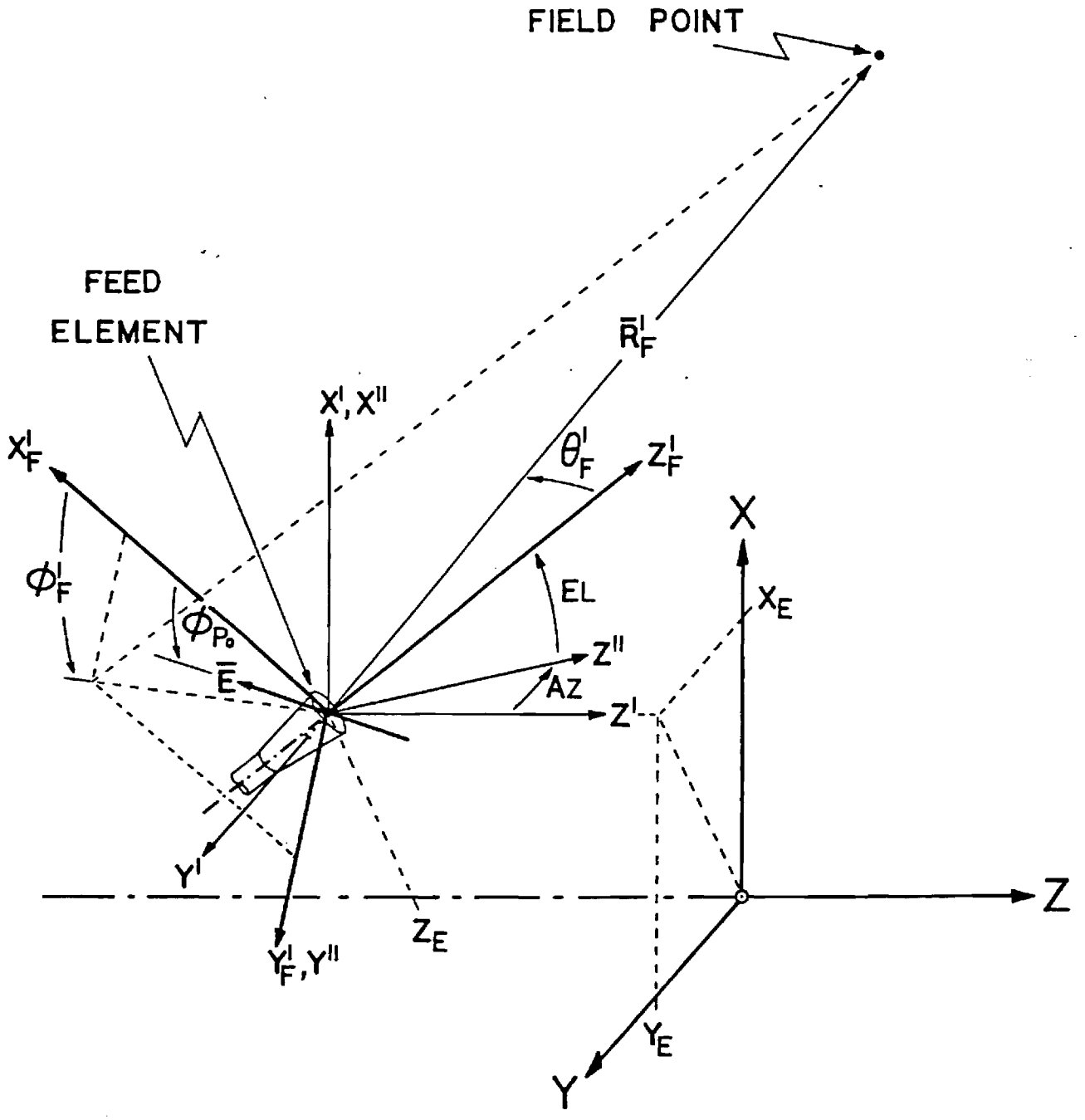


Fig. 3.1 - Feed element geometry.

$$X_F' = X'' \cos EL - Z'' \sin EL \qquad (3.3\,a)$$
$$Y_F' = Y'' \qquad (3.3\,b)$$
$$Z_F' = X'' \sin EL + Z'' \cos EL \qquad (3.3\,c)$$

$$R_F' = \sqrt{X_F'^2 + Y_F'^2 + Z_F'^2} \qquad (3.4\,a)$$

$$\theta_F' = \arccos\left[\frac{\sqrt{X_F'^2 + Y_F'^2}}{R_F}\right] \qquad (3.4\,b)$$

$$\phi_F' = \arctan\left[Y_F'/X_F'\right] \qquad (3.4\,c)$$

And their inverses

$$X = X' + X_E \qquad (3.5\,a)$$
$$Y = Y' + Y_E \qquad (3.5\,b)$$
$$Z = Z' + Z_E \qquad (3.5\,c)$$

$$X' = X'' \qquad (3.6\,a)$$
$$Y' = Y'' \cos Az - Z'' \sin Az \qquad (3.6\,b)$$
$$Z' = Y'' \sin Az + Z'' \cos Az \qquad (3.6\,c)$$

$$X'' = X_F' \cos EL + Z_F' \sin EL \qquad (3.7\,a)$$
$$Y'' = Y_F' \qquad (3.7\,b)$$
$$Z'' = -X_F' \sin EL + Z_F' \cos EL \qquad (3.7\,c)$$

$$X_F' = \rho_F' \sin\theta_F' \cos\phi_F' \qquad (3.8\,a)$$
$$Y_F' = \rho_F' \sin\theta_F' \sin\phi_F' \qquad (3.8\,b)$$
$$Z_F' = \rho_F' \cos\theta_F' \qquad (3.8\,c)$$

Starting with the field point coordinates X, Y, Z (in the principal system), eqs. (3.1) - (3.4) allow the coordinates R_F', θ_F', ϕ_F' of P to be obtained.

Each feed element is assumed to radiate electric and magnetic fields, $\bar{E}_F$ and $\bar{H}_F$ respectively given by the equations:

$$\bar{E}_F = EEXCIT \frac{e^{-j\beta R'_F}}{R'_F} \left\{ \cos^E \theta'_F \cos(\phi'_F - \phi_{P_0}) \hat{\theta}'_F - \cos^H \theta'_F \sin(\phi'_F - \phi_{P_0}) \hat{\phi}'_F \right\}, \quad (3.9a)$$

$$\bar{H}_F = \frac{1}{\eta} \hat{R}'_F \times \bar{E}_F, \quad (3.9b)$$

FOR $0° \le \theta'_F \le 90°$, AND $\bar{E}_F = 0$ FOR $90° < \theta'_F < 180°$.

These equations correspond to a perfectly linearly polarized, according to Ludwig's third polarization definition [3.1], spherical wave point source feed. The pattern shape can be adjusted independently in both principal planes through the exponents E and H, due to this fact this feed is called a raised – cosine – type feed element. EEXCIT is the element complex excitation coefficient. Although in here we are assuming that E, H, and ϕ_{P_0} remain constant for all array elements, EEXCIT can be adjusted individualy for each one of them. The raised-cosine-type pattern is analytically very convenient since it is fully specified by few parametens, but it shovld be noted howeven that it also provides a good first approximation for practical feed patterns.

In Eqs. (3.9) the $\hat{R}'_F$ components of the fields are zero and also the wave is TEM. due to this they are a far-field representation. In the dual-reflector antenna, the subreflecton will be in the near field of each array element, in many cases. Even then Eqs. (3.9) are good feed pattern approximations, provided that the proper E and H exponents are used.

Now, returning to Eqs. (3.1)-(3.8), once the field point coordinates R'_F, θ'_F, and ϕ'_F are known, Eqs. (3.9) give the electric and magnetic fields at P. Physically

speaking Eqs. (3.1)-(3.8) simply mean that each feed is mounted on an independent elevation over azimuth positioner. A_z and E_L are the azimuth and elevation angles, respectively ($A_z = E_L = 0$ for a feed aimed towards the $\hat{z}$ direction.

The feed field components given by Eqs. (3.9) are the feed spherical system ones. For posterior use it is convenient to express them in the principal cartesian system X, Y, Z. Calling $\bar{E}_s$ and $\bar{H}_s$ the feed field in the X, Y, Z principal system, it can be written:

$$\bar{E}_s = (\bar{E}_F \cdot \hat{x})\hat{x} + (\bar{E}_F \cdot \hat{Y})\hat{Y} + (\bar{E}_F \cdot \hat{z})\hat{z}, \qquad (3.10a)$$

$$\bar{H}_s = (\bar{H}_F \cdot \hat{x})\hat{x} + (\bar{H}_F \cdot \hat{Y})\hat{Y} + (\bar{H}_F \cdot \hat{z})\hat{z}. \qquad (3.10b)$$

Where the unit vectors $\hat{x}, \hat{Y}$, and $\hat{z}$ (of the principal system) are obtained using Eqs. (3.5)-(3.8). Neglecting the non existent $\hat{R}_F$ components, they are given by:

$$\hat{X} = \hat{\theta}'_F\left[\cos\theta'_F \cos\phi'_F \cos E_L - \sin\theta'_F \sin E_L\right] - \hat{\phi}'_F\left[\sin\phi'_F \cos E_L\right], \qquad (3.11a)$$

$$\hat{Y} = \hat{\theta}'_F\left[(\cos\theta'_F \cos\phi'_F \sin E_L + \sin\theta'_F \cos E_L)\sin A_z + \cos\theta'_F \sin\phi'_F \cos A_z\right] + \hat{\phi}'_F\left[\cos\phi'_F \cos A_z - \sin\phi'_F \sin E_L \sin A_z\right], \qquad (3.11b)$$

$$\hat{Z} = \hat{\theta}'_F\left[-(\cos\theta'_F \cos\phi'_F \sin E_L + \sin\theta'_F \cos E_L)\cos A_z + \cos\theta'_F \sin\phi'_F \sin A_z\right] + \hat{\phi}'_F\left[\cos\phi'_F \sin A_z + \sin\theta'_F \sin E_L \cos A_z\right]. \qquad (3.11c)$$

By adding up all the cartesian components produced by each array element the total array field produced at the point P can be obtained. In this way the field incident on the subreflector surface can be determined. Note that, even though the point P is assumed to be in the far-field of each array element, it is in general in the near-field of the array, and this is fully taken into account.

4 - Subreflecton Scattering

This section addresses the computation of the scattering, of the array feed fields, by the subreflector. Two methods are consided, the Geometrical Theory of Diffraction (GTD) and the Physical Optics (PO) one. Each method has its own advantages and short comings, and in many ways they complement each other; due to this fact both are explored in this work. In the next few pages the application of both methods to the subreflector scattering computation is presented in detail.

4.1 - Geometrical Theory of Diffraction

Consider Fig. 4.1 showing a source point S, a subreflector with arbitrary surface and rim shapes, and a field point P. At S a spherical wave point source is located (e.g. one of the array feed elements), and our problem in here is to determine the field scattered by the subreflector at P (P can be, for instance, a point on the main-reflector surface). According to the GTD principles and validity conditions [4.1], the field at P is basicaly due to two types of rays, reflected and diffracted, which emanate from reflection and diffraction points, respectively. These rays obey all ray propagation laws, and their corresponding electromagnetic field values can be obtained from the antenna geometry and illumination using well documented reflection and diffraction tensor coefficients [1.3]. The reflection and diffraction points are located on the subreflector surface and rim, respectively, and they are determined by the generalized Fermat's principle [4.1].

The computation of the field at P can be divided in four well defined parts, reflection and diffraction points determination, and computation of the reflected and diffracted ray fields. We consider these four parts separately in the next pages, and obtain the subreflector scattering by adding the field due to the two GTD-type rays. By repeating this procedure for each array element, and adding the scattering

contribution of each, the complete sub-reflector scattering is determined.

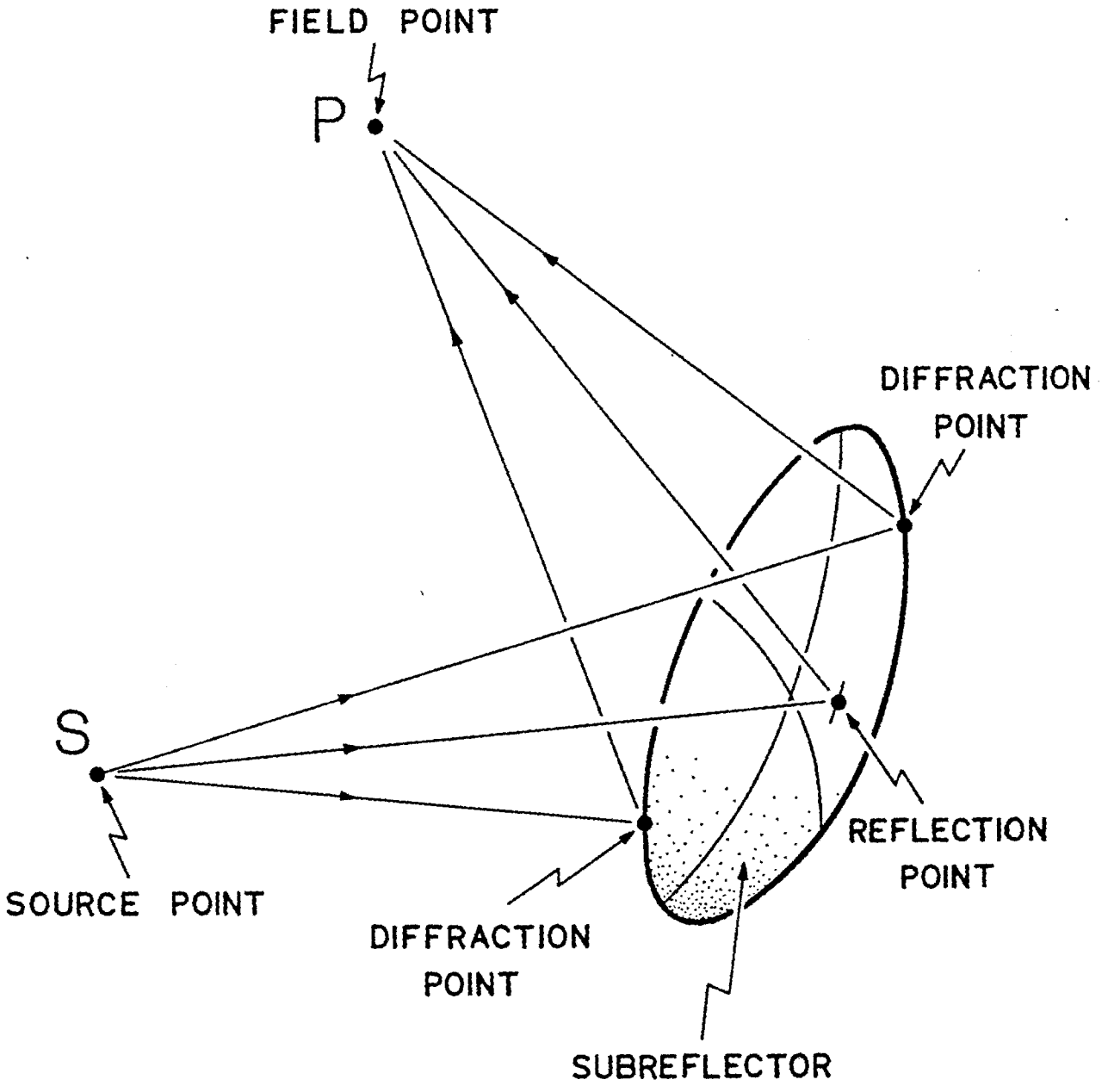


Fig. 4.1 - Subreflecton scattering mechanism according to the geometrical theory of diffraction.

4.1.1 - Reflection-points Determination

Let's assume that the subreflector of Fig. 4.1 is described, in the principal system cartesian coordinates, by the equation

$$z = h(x,y), \qquad (4.1.1)$$

where h is a known function of the variable x and y. Our objective here is to determin the coordinates X, Y, and X of the reflection point R, where a ray going from S to P, bounces off at the subreflecto The geometry for the reflection-point determination is shown in Fig. 4.2.

The distances l_s and l_p (between the points S and R, and R and P, respectively) are given by

$$l_s = \sqrt{(X-X_s)^2 + (Y-Y_s)^2 + (h(x,y) - Z_s)^2}, \qquad (4.1.2a)$$

$$l_p = \sqrt{(X-X_p)^2 + (Y-Y_p)^2 + (h(x,y) - Z_p)^2}. \qquad (4.1.2b)$$

According to Fermat's principle [4.2], if R is a reflection point for a ray going from S to P (through a reflection at the subreflector), then the total path length $l = l_s + l_p$ must be stationary for a variation of R on the subreflector surface. Writing this in mathematical form one have the conditions

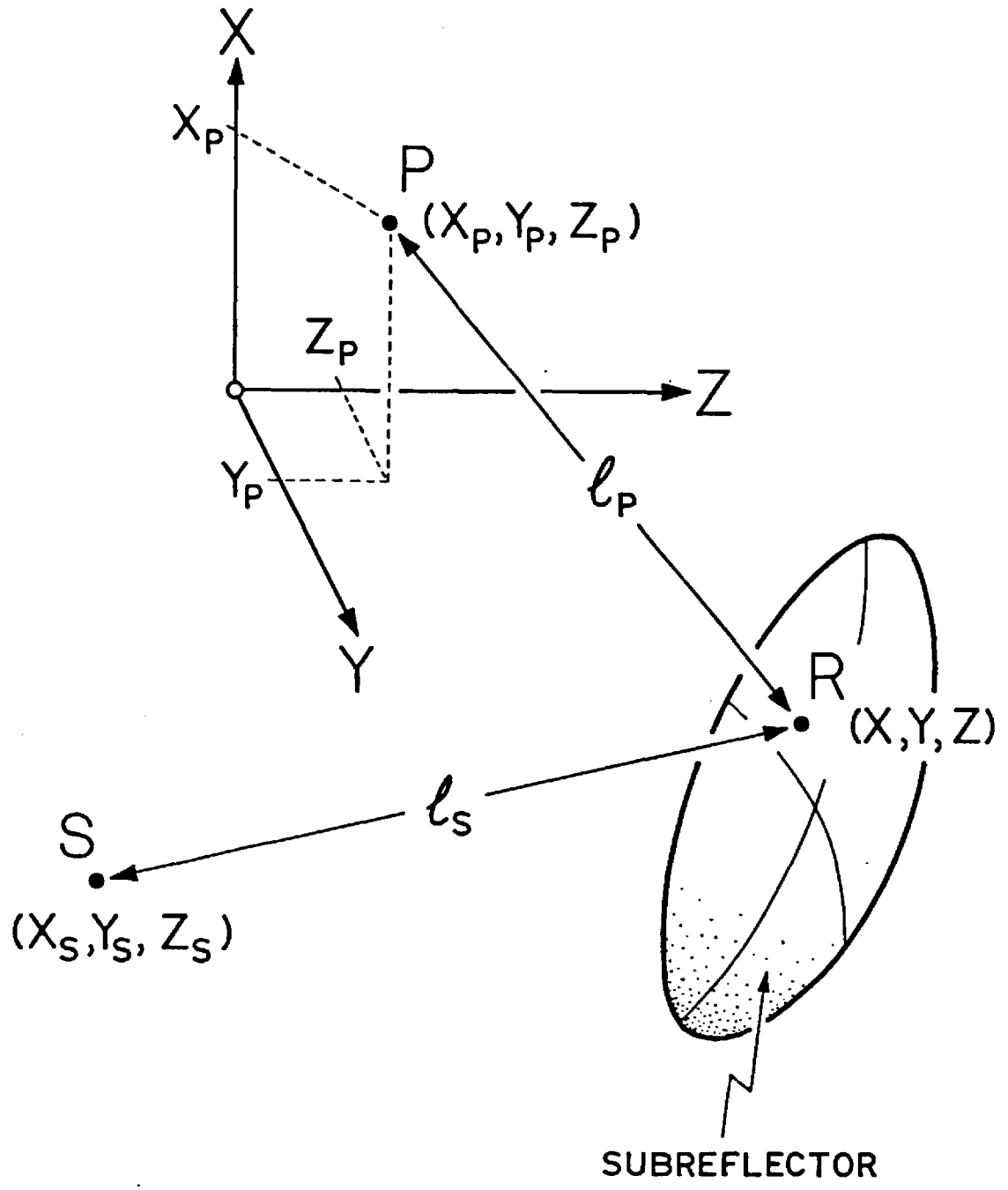


Fig. 4.2 - Geometry for the subreflecton reflection points determination.

$$\frac{\partial l}{\partial x} = \frac{\partial l_s}{\partial x} + \frac{\partial l_p}{\partial x} = 0 , \qquad (4.1.3a)$$

$$\frac{\partial l}{\partial Y} = \frac{\partial l_s}{\partial Y} + \frac{\partial l_p}{\partial Y} = 0 . \qquad (4.1.3b)$$

Since, from Eqs. (4.1.2),

$$\frac{\partial l_s}{\partial x} = \frac{(x-x_s) + (h-z_s)\,\partial h/\partial x}{l_s} , \qquad (4.1.4a)$$

$$\frac{\partial l_s}{\partial Y} = \frac{(Y-Y_s) + (h-z_s)\,\partial h/\partial Y}{l_s} , \qquad (4.1.4b)$$

$$\frac{\partial l_p}{\partial x} = \frac{(x-x_p) + (h-z_p)\,\partial h/\partial x}{l_p} , \qquad (4.1.4c)$$

$$\frac{\partial l_p}{\partial Y} = \frac{(Y-Y_p) + (h-z_p)\,\partial h/\partial Y}{l_p} , \qquad (4.1.4d)$$

one can then write Eqs. (4.1.3) as

$$F(X,Y) = \left[(x-x_s) + (h-z_s)\frac{\partial h}{\partial x}\right] l_p + \left[(x-x_p) + (h-z_p)\frac{\partial h}{\partial x}\right] l_s = 0 , \quad (4.1.5$$

$$G(X,Y) = \left[(Y-Y_s) + (h-z_s)\frac{\partial h}{\partial Y}\right] l_p + \left[(Y-Y_p) + (h-z_p)\frac{\partial h}{\partial Y}\right] l_s = 0 . \quad (4.1.$$

The above Eqs. (4.1.5) form a non-linear system of two equations with two unknowns, X and Y. Its solutions gives the desired reflection points coordinates Note that, to simplify notation, we have abbreviated the function $h(X,Y)$ by h.

In this work we have solved the system given by Eqs. (4.1.5) using the

two dimensional Newton's method [4.3]. This iterative procedure starts with a guess X_m and Y_m for the reflection point coordinates and then determines a better approximation X_{m+1} an Y_{m+1} through the equations

$$X_{m+1} = X_m + \delta , \tag{4.1.6a}$$

$$Y_{m+1} = Y_m + \epsilon , \tag{4.1.6b}$$

where

$$\delta = \frac{GF_Y - FG_Y}{F_X G_Y - F_Y G_X} , \tag{4.1.7a}$$

$$\epsilon = \frac{FG_X - GF_X}{F_X G_Y - F_Y G_X} , \tag{4.1.7b}$$

And F_X, F_Y, G_X, and G_Y are obtained from Eqs. (4.1.5) as

$$F_X = \frac{\partial F(X,Y)}{\partial X} = \left[1 + \left(\frac{\partial h}{\partial x}\right)^2\right] l + \left[(h-z_s) l_p + (h-z_p) l_s\right] \frac{\partial^2 h}{\partial x^2} +$$

$$+ \left[(X-X_s) + (h-z_s)\frac{\partial h}{\partial x}\right] \frac{\partial l_p}{\partial x} + \left[(X-X_p) + (h-z_p)\frac{\partial h}{\partial x}\right] \frac{\partial l_s}{\partial x} , \tag{4.1.8a}$$

$$F_Y = \frac{\partial F(X,Y)}{\partial Y} = \left(\frac{\partial h}{\partial x}\right)\left(\frac{\partial h}{\partial Y}\right) l + \left[(h-z_s) l_p + (h-z_p) l_s\right] \frac{\partial^2 h}{\partial x \partial Y} +$$

$$+ \left[(X-X_s) + (h-z_s)\frac{\partial h}{\partial x}\right] \frac{\partial l_p}{\partial Y} + \left[(X-X_p) + (h-z_p)\frac{\partial h}{\partial x}\right] \frac{\partial l_s}{\partial Y} , \tag{4.1.8b}$$

$$G_x = \frac{\partial G(X,Y)}{\partial X} = \left(\frac{\partial h}{\partial x}\right)\left(\frac{\partial h}{\partial Y}\right) l + \left[(h-z_s)l_p + (h-z_p)l_s\right] \frac{\partial^2 h}{\partial x \partial Y} +$$

$$+ \left[(Y-Y_s) + (h-z_s)\frac{\partial h}{\partial Y}\right]\frac{\partial l_p}{\partial x} + \left[(Y-Y_p) + (h-z_p)\frac{\partial h}{\partial Y}\right]\frac{\partial l_s}{\partial x}, \quad (4.1.8c)$$

$$G_Y = \frac{\partial G(X,Y)}{\partial Y} = \left[1 + \left(\frac{\partial h}{\partial Y}\right)^2\right] l + \left[(h-z_s)l_p + (h-z_p)l_s\right] \frac{\partial^2 h}{\partial x^2} +$$

$$+ \left[(Y-Y_s) + (h-z_s)\frac{\partial h}{\partial Y}\right]\frac{\partial l_p}{\partial Y} + \left[(Y-Y_p) + (h-z_p)\frac{\partial h}{\partial Y}\right]\frac{\partial l_s}{\partial Y}. \quad (4.1.8d)$$

Starting with a good initial guess for X and Y (X_m and Y_m with $m=1$), and using Eqs. (4.1.8) repeatedely, the reflection point position can be obtained with sufficient engineering accuracy after only a few iteractions. This is due to the fact that Newton's method is guaranteed to converge (with quadratic rate) to the solution of Eqs. (4.1.8) provided that one does indeed exists and that the initial guess is sufficiently close to the solution [4.3]. In this work one has to determine the reflection point positions for several different S and P locations. We then arbitrarily used, as a starting guess, the values $X=Y=0$ for the first pair S and P. Once the corresponding reflection point position was obtained, its coordinates were then used as the starting guess for the next source and observation point pairs, and so on. This procedure proved to be very effective since each new S and P pairs were almost always near the previous ones and hence only very few iteractions were needed to obtain the new reflection point position. In the cases where the

NEW S AND P PAIRS WERE NOT CLOSE TO THE PREVIOUS ONES, NEWTON'S METHOD HAD ONLY TO ITERATE A LITTLE MORE - WE COULD HAVE USED X=Y=0 FOR THESE CASES, BUT IN PRINCIPLE THERE IS NO REASON TO BELIVE THAT THIS POSITION PROVIDES A BETTER STARTING POINT.

FOUR TESTS WERE USED TO STOP THE NEWTON'S METHOD ITERATIONS. THE FIRST IS SIMPLY LIMITING THE NUMBER OF ITERATIONS TO 10 - EXPERIENCE SHOWS THAT THIS NUMBER IS MORE THAN SUFFICIENT TO OBTAIN THE REFLECTION POINTS POSITIONS WITH THE ACCURACY REQUIRED FOR THIS WORK. THE SECOND STOPS THE ITERATIONS IF THE CONDITION

$$\sqrt{(X_{m+1}-X_m)^2+(Y_{m+1}-Y_m)^2+(Z_{m+1}-Z_m)^2} > \text{DIVERG}$$

IS SATISFIED (WITH DIVERG EQUAL TO A LARGE NUMBER); THIS TEST PREVENTS THE CONTINUATION OF A DIVERGENT PROCESS. THE THIRD TEST IS TRIGGERED BY THE SURFACE EQUATION [Eq. (4.1.1)]; FOR A GIVEN PAIR X, Y IT MAY BE IMPOSSIBLE TO DETERMINE THE VALUE OF Z, IN WHICH CASE THE ITERATIONS CAN'T PROCEED (THIS MAY HAPPEN, FOR EXAMPLE, IN AN ELLIPSOIDAL SUBREFLECTOR SINCE THE SURFACE IS DEFINED ONLY FOR X AND Y IN A CERTAIN DOMAIN). THE FORTH AND LAST TEST VERIFIES THAT A REFLECTION POINT HAS BEEN FOUND, AND THAT ITS COORDINATES HAVE BEEN DETERMINED WITH SUFFICIENT ACCURACY. IT STOPS THE ITERATIONS WHENEVER

$$\sqrt{(X_{m+1}-X_m)^2+(Y_{m+1}-Y_m)^2+(Z_{m+1}-Z_m)^2} \leq 0.0012,$$

$\partial l/\partial x \leq 0.001$, and $\partial l/\partial y \leq 0.001$. The square root condition guarantees that, from one iteration to the next, the reflection point position has not moved by more than 0.001 wavelengths. The derivative checks guarantees that one indeed have a stationary path length.

In the event that the iterations were stoped by any of the first three above mentioned tests, a new search (with the same S and P positions) is started using for guesses half the valves of X and Y last obtained. This new search is started at most ten times before the no-reflection-point-available fact is accepted. This procedure was included to provide a redundancy against any bad starting guesses.

Once Newton's method yields a reflection point its coordinates are tested to verify if it is inside the subreflector rim. This test is required since Eq. (4.1.1) does not provide information regarding the actual subreflector size. Only if this last check is satisfied one have found a physical reflection point.

Throughout this section we have assumed that the subreflector surface is described by an equation of the type given by Eq. (4.1.1). Since in Sec. 2 we only presented the subreflector surface in spherical coordinates [i.e. by Eq. (2.7)] we now present the corresponding cartesian form for future reference.

Using the cartesian system X_s, Y_s, Z_s defined by

$$X_s = X\cos\beta - Z\sin\beta \,, \tag{4.1.9a}$$

$$Y_s = Y \,, \tag{4.1.9b}$$

$$Z_s = X\sin\beta + Z\cos\beta \,, \tag{4.1.9c}$$

the subreflector equation, Eq. (2.7) can be written into the well known form

$$Z_s = +\sqrt{\frac{X_s^2+Y_s^2}{e^2-1} + \left(\frac{c}{e}\right)^2} - c \,. \tag{4.1.10}$$

The X_s, Y_s, Z_s system is the principal system rotated by an angle β about the Y axis (the subreflector axis coincides with the z-axis).

Substituting Eqs. (4.1.9) into Eq. (4.1.10) one obtains

$$Az^2 + Bz + C = 0 \,, \tag{4.1.11}$$

where

$$A = e^2\cos^2\beta - 1 \,, \tag{4.1.12a}$$

$$B = 2\cos\beta\,[x e^2 \sin\beta + c(e^2-1)] \,, \tag{4.1.12b}$$

$$C = (e^2-1)\left[(x\sin\beta + c)^2 - (c/e)^2\right] - x^2\cos^2\beta - y^2 \,. \tag{4.1.12c}$$

Equation (4.1.11) can then be solved to yield the form stipulated by Eq. (4.1.1).

As a final remark it should be mentioned that, for a general geometry, Eqs. (4.1.5) may have several solutions (eg. several reflection points yielding rays going from

point S to point P). However, for the systems considered in this work, we have verified that normally only one reflection point is present on the subreflector. Due to this fact we have not considered in here the more involved problem of searching for multiple reflection points

4.1.2 - Diffraction-points Determination

Consider once more Fig. 4.1, as shown there the diffraction points occur in the periphery of the subreflecton. Due to this fact, throughout this section we are concerned only with points laying on the curve that defines the subreflector rim. Since the rim is in general a tri-dimensional curve, it is convenient to assume that its X, Y, and Z coordinates are given in parametric for through the equations

$$X = X(t), \tag{4.1.13a}$$

$$Y = Y(t), \tag{4.1.13b}$$

$$Z = Z(t), \tag{4.1.13c}$$

where t is an arbitrary parameter. Since the rim is a closed curve we can assume, without loss in generality, that as t goes from 1 to a certain value t_M the complete rim is scanned only once [note that $X(1) = X(t_M)$, $Y(1) = Y(t_M)$ and $Z(1) = Z(t_M)$].

The parametric rim description according to Eqs. (4.1.13) is convenient because any rim point is specified by a single parameter t. This greately simplifies the search for diffraction points and the computation of the rim geometrical parameters, and also allows for complete generality in the rim shape.

Figure 4.3 presents the geometry for the

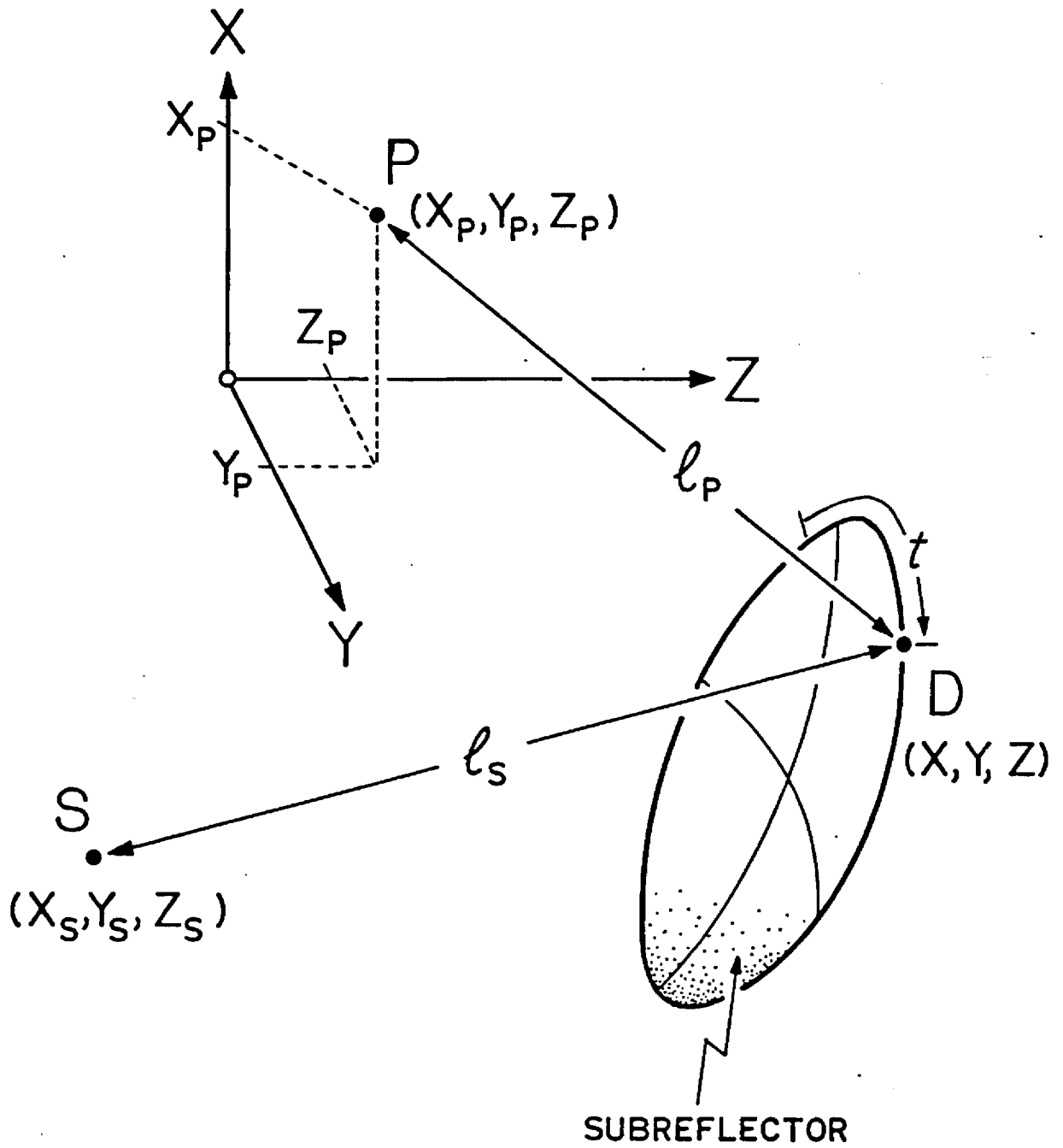


Fig. 4.3 - Geometry for the subreflector diffraction points determination.

subreflector diffraction-points determination. Given a source and a field point locations S and P respectively, and an arbitrary subreflector rim defined according to Eqs. (4.1.13), in this section we present a method to find the points D, where a ray emanating from S diffracts to illuminate the point P. It can be easily seen that, for any smooth subreflector rim, an even number of diffraction points always exist (two being the minimum number), provided that the concept of diffraction points is meaningful – we then have always to determine a minimum of two diffraction points.

Consider now the total path-length l from point S to point P, going through the point D. According to Fig. 4.3 we can write that $l = l_s + l_p$ where

$$l_s = \sqrt{(X(t)-X_s)^2 + (Y(t)-Y_s)^2 + (Z(t)-Z_s)^2}\,, \qquad (4.1.14a)$$

$$l_p = \sqrt{(X(t)-X_p)^2 + (Y(t)-Y_p^2) + (Z(t)-Z_p)^2}\,. \qquad (4.1.14b)$$

Using the generalized Fermat's principle [4.1] one know that, if D is a diffraction point, its associated path-length l must be stationary with respect to a displacement of D along the subreflector rim. In mathematical form this statement is equivalent to saying that, at a diffraction point,

$$\frac{dl}{dt} = \frac{dl_s}{dt} + \frac{dl_p}{dt} = 0\,. \qquad (4.1.15)$$

Since [from Eqs. (4.1.14)]

$$\frac{dl_s}{dt} = \frac{1}{l_s}\left[(x-x_s)\frac{dx}{dt} + (y-y_s)\frac{dy}{dt} + (z-z_s)\frac{dz}{dt}\right], \qquad (4.1.16a)$$

$$\frac{dl_p}{dt} = \frac{1}{l_p}\left[(x-x_p)\frac{dx}{dt} + (y-y_p)\frac{dy}{dt} + (z-z_p)\frac{dz}{dt}\right], \qquad (4.1.16b)$$

one can then write Eq. (4.1.15) as

$$F(t) = l_p\left[(x-x_s)\frac{dx}{dt} + (y-y_s)\frac{dy}{dt} + (z-z_s)\frac{dz}{dt}\right] +$$

$$+ l_s\left[(x-x_p)\frac{dx}{dt} + (y-y_p)\frac{dy}{dt} + (z-z_p)\frac{dz}{dt}\right] \qquad (4.1.17)$$

This nonlinear equation is the diffraction points equation; its solutions correspond to the parametric positions t of all subreflector diffraction points. Note that, for notation simplicity, the rim coordinates $x(t)$, $y(t)$, and $z(t)$ are being abreviated by x, y, and z.

In this work we have solved Eq. (4.1.17) numerically using a step-search combined with the one-dimensional version of Newton's method [4.3]. We started our step-search from a point t corresponding to a diffraction point. We then walked along the subreflector rim, in steps of about one wavelength, checking for zero-crossings of the function $F(t)$ [given by Eq. (4.1.17)]. Once a zero-crossing was detected Newton's method was used to refine the t-value yielded by the step-search. By scanning the subreflector rim completely we then assured that all reflection

points were obtained.

Newton's method refines the t values yielded by the step-search according to the iterative prescription [4.3]

$$t_{m+1} = t_m + \delta , \tag{4.1.18}$$

where

$$\delta = - \left. \frac{F(t)}{dF(t)/dt} \right|_{t=t_m} . \tag{4.1.19}$$

In Eq (4.1.18) t_{m+1} is the refined value of t_m, the subscript m referring to the mth iteration. The expression for $dF(t)/dt$, obtained from Eq. (4.1.17), is given below for future reference.

$$\frac{dF(t)}{dt} = \left[\left(\frac{dl_s}{dt}\right)\left(\frac{dl_p}{dt}\right) + \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 \right] l +$$

$$+ l_p \left[(x-x_s)\frac{d^2x}{dt^2} + (y-y_s)\frac{d^2y}{dt^2} + (z-z_p)\frac{d^2z}{dt^2} \right] +$$

$$+ l_s \left[(x-x_p)\frac{d^2x}{dt^2} + (y-y_p)\frac{d^2y}{dt^2} + (z-z_p)\frac{d^2z}{dt^2} \right] . \tag{4.1.20}$$

As mentioned above, the step-search procedure was initiated at a diffraction point. This first diffraction point was obtained through Newton's method using repeated guesses for t until a diffraction point was found (very seldom more than one guess was required). At most ten t-guesses (equispaced over the rim) were allowed before deciding that no well defined diffraction points existed.

Three tests were used to monitor the operation of Newton's method. The first simply sets the maximum number of iterations to 10. The second stops the iterations when the value of t becomes greater than t_M [see paragraph following Eqs. (4.1.13) for t_M]. And the third stops the iterations whenever

$$\sqrt{(x_{m+1}-x_m)^2+(y_{m+1}-y_m)^2+(z_{m+1}-z_m)^2} \leq 0.001\,\lambda,$$

and

$$\frac{dl}{ds} = \frac{dl}{dt}\times\frac{dt}{ds} \leq 0.001;$$

where ds is the differential distance element along the subreflector rim, given by the formula

$$\frac{ds}{dt} = \sqrt{\left(\frac{dx}{dt}\right)^2+\left(\frac{dy}{dt}\right)^2+\left(\frac{dz}{dt}\right)^2}.$$

The sole purpose of tests one and two is to prevent the algorithm from running wild in the possible event of an ill posed geometry - this safeguard sometimes operates when the rim does not have well defined diffraction points. And test three guarantees that the diffraction points have been determined with sufficient engineering accuracy, and that they indeed correspond to stationary path length conditions.

Throughout this section the subreflector has been assumed defined in parametric form according to Eqs. (4.1.13). These

PARAMETRIC EQUATIONS WERE OBTAINED BY DETERMINING THE X, Y, AND Z COORDINATES OF THE RIM AT POINTS SEQUENTIALLY SPACED BY ABOUT ONE WAVELENGTH AND THEN ASSIGNING THE VALUE $t=1$ FOR THE FIRST COORDINATE TRIO, $t=2$ FOR THE SECOND, AND SO ONE UNTIL THE COMPLETE RIM WAS COVERED. NEXT THREE CUBIC SPLINES [4.3] WERE FITED THROUGH THESE RIM COORDINATES ONE FOR EACH COORDINATE. THESE SPLINES THEN YIELDED THE REQUIRED PARAMETRIC FORM FOR THE RIM COORDINATES, IN A PROCEDURE APLICABLE TO ANY RIM SHAPE. AS IS APARENT FROM FIG. 4.3 THE RIM CAN BE SPECIFIED IN A CLOCKWISE OR COUNTER CLOCKWISE FASHION AS THE t VALUES INCREASE. THE CHOICE IS ARBITRARY BUT ONCE MADE MUST BE MANTAINED SINCE THE COMPUTATION OF THE DIFFRACTED FIELDS IS DEPENDENT ON IT. OUR RIM WAS DEFINED IN A CLOCKWISE SENSE WHEN OBSERVED FROM THE SUBREFLECTOR SURFACE SIDE INTO WHICH THE SURFACE UNIT NORMAL IS POINTING. ALSO, THE X, Y, AND Z COORDINATES WERE OBTAINED USING EQ. (2.7).

4.1.3 - Reflected-Field Evaluation

In Sec. 4.1.1 we presented a procedure to determine the reflection point coordinates X, Y, and Z, for a ray going from a source point S to a field point P (see Fig. 4.2). Assuming then that we now have the reflection point position, in this section we evaluate the complex vector field reflected by the subreflector at point P.

According to the ray picture of the electromagnetic energy propagation, reflection is a localized phenomenon governed by the surface characteristics in the vicinity of the reflection point. Let's consider then the arbitrary subreflector surface element shown in Fig. 4.4. Differential geometry shows that any analytical surface can be approximated, in the neighborhood of a point, by a second degree surface [4.4], [4.5] — this is equivalent to expanding the surface equation in a tri-dimensional Taylor's series and retaining only terms up to the second power. The intersection of any plane containing the surface unit normal $\hat{N}_s$, with the surface can then be approximated by a circle. In particular the planes containing the unit vectors $\hat{U}_1$ and $\hat{U}_2$ yield circles with maximum and minimum radii R_1 and R_2, respectively. R_1, R_2, $\hat{U}_1$ an $\hat{U}_2$ completely characterize the surface in the neighborhood of the X, Y, Z point — R_1 and R_2 are called the principal radii of curvature of the surface, and $\hat{U}_1$ and $\hat{U}_2$ their associated principal directions. Convenient formulas for determining R_1, R_2, $\hat{U}_1$ and $\hat{U}_2$

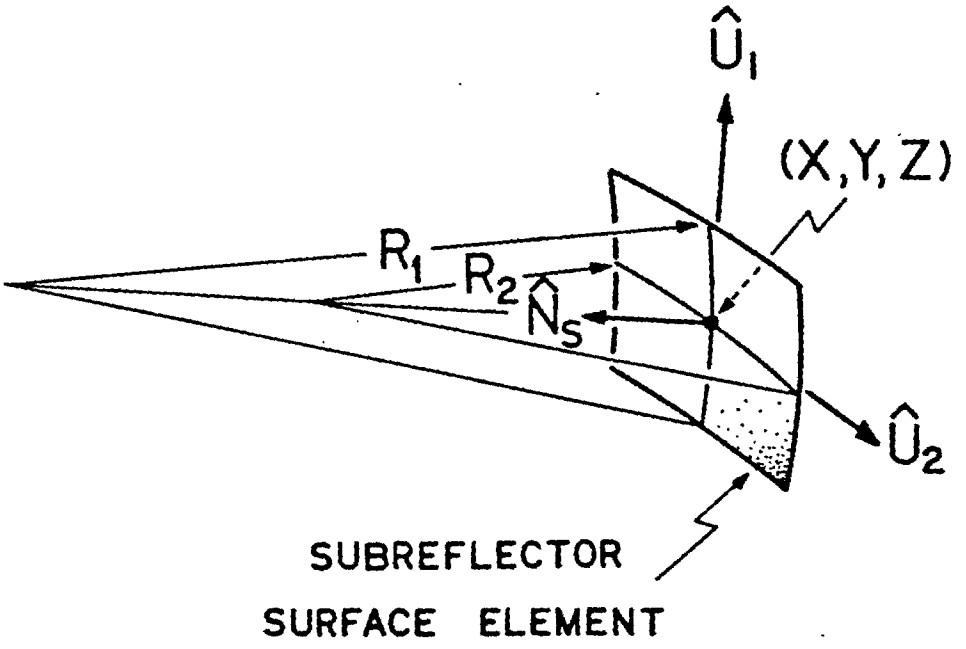


Fig. 4.4 - Approximation of the subreflector surface element by a second degree surface.

Are available in [4.4]; they are given below for completeness.

According to Eq. (4.1.1) the subreflector surface is described by

$$z = h(x, y).$$

The principal radii of curvature are then given by

$$R_1 = 1/\kappa_1, \tag{4.1.22a}$$

$$R_2 = 1/\kappa_2, \tag{4.1.22b}$$

where the curvatures κ_1 and κ_2 are equal to

$$\begin{matrix}\kappa_1 \\ \kappa_2\end{matrix} = \kappa_m \pm \sqrt{\kappa_m^2 - \kappa_G}. \tag{4.1.23}$$

In the above equation κ_m and κ_G are the mean and Gaussian curvatures of the surface, respectively, given by

$$\kappa_m = \frac{E'g' - 2f'F' + e'G'}{2D'}, \tag{4.1.24a}$$

$$\kappa_G = \frac{e'g' - f'^2}{D'}, \tag{4.1.24b}$$

with

$$D' = 1 + \left(\frac{\partial h}{\partial x}\right)^2 + \left(\frac{\partial h}{\partial y}\right)^2, \tag{4.1.25a}$$

$$E' = 1 + \left(\frac{\partial h}{\partial x}\right)^2, \tag{4.1.25b}$$

$$F' = \left(\frac{\partial h}{\partial x}\right)\left(\frac{\partial h}{\partial Y}\right), \tag{4.1.25c}$$

$$G' = 1 + \left(\frac{\partial h}{\partial x}\right)^2, \tag{4.1.25d}$$

$$e' = \Delta' \frac{\partial^2 h}{\partial x^2}, \tag{4.1.25e}$$

$$f' = \Delta' \frac{\partial^2 h}{\partial Y^2}, \tag{4.1.25f}$$

$$g' = \Delta' \frac{\partial^2 h}{\partial Y^2}, \tag{4.1.25g}$$

$$\Delta' = \frac{1}{s'\sqrt{D'}}. \tag{4.1.25h}$$

In the above equations we have abbreviated $h(x,Y)$ by h.

The s' of Eq. (4.1.25h) takes care of the sign ambiguity present in the $\sqrt{D'}$. Its value is either $+1$ or -1, and it is chosen according to the desired direction of the unit normal $\hat{N}_s$, given by

$$\hat{N}_s = s' \frac{\frac{\partial \bar{r}_s}{\partial x} \times \frac{\partial \bar{r}_s}{\partial Y}}{\left|\frac{\partial \bar{r}_s}{\partial x} \times \frac{\partial \bar{r}_s}{\partial Y}\right|}, \tag{4.1.26}$$

where $\bar{r}_s$ is a vector from the principal coordinate system origen to the subreflector surface point, more specifically

$$\bar{r}_s = x\hat{x} + Y\hat{Y} + h\hat{z}. \tag{4.1.27}$$

From the above equation one obtain

$$\frac{\partial \bar{n}_s}{\partial x} = \hat{x} + \hat{z}\frac{\partial h}{\partial x}, \qquad (4.1.28a)$$

$$\frac{\partial \bar{n}_s}{\partial y} = \hat{y} + \hat{z}\frac{\partial h}{\partial y}, \qquad (4.1.28b)$$

And then

$$\frac{\partial \bar{n}_s}{\partial x} \times \frac{\partial \bar{n}_s}{\partial y} = \hat{x}\frac{\partial h}{\partial x} + \hat{y}\frac{\partial h}{\partial y} - \hat{z}. \qquad (4.1.29)$$

Since $\hat{N}_s$ should point into the source side of the subreflector surface our choice for s' must be

$$s' = +1. \qquad (4.1.30)$$

The vectors $\bar{U}_1'$ and $\bar{U}_2'$ is the principal directions are given by

$$\bar{U}_1' = \frac{\partial \bar{n}_s}{\partial x} + \alpha' \frac{\partial \bar{n}_s}{\partial y}, \qquad (4.1.31a)$$

$$\bar{U}_2' = \beta \frac{\partial \bar{n}_s}{\partial x} + \frac{\partial \bar{n}_s}{\partial y}, \qquad (4.1.31b)$$

where

$$\alpha' = \frac{f' - \kappa_1 F'}{\kappa_1 G' - g'}, \qquad (4.1.32a)$$

$$\beta' = \frac{g' - \kappa_2 G'}{\kappa_2 F' - f'}. \qquad (4.1.32b)$$

Substituting Eqs. (4.1.28) and (4.1.32) in Eqs. (4.1.31), and eliminating the denominators to avoid division by zero problems, one arrives at a new pair of vectors pointing along the principal directions,

$$\bar{U}_1 = (\kappa_1 G' - g')\hat{X} + (f' - \kappa_1 F')\hat{Y} + \left[(\kappa_1 G' - g')\frac{\partial h}{\partial x} + (f' - \kappa_1 F')\frac{\partial h}{\partial y}\right]\hat{Z}, \quad (4.1.33a)$$

$$\bar{U}_2 = (g' - \kappa_2 G')\hat{X} + (\kappa_2 F' - f')\hat{Y} + \left[(g' - \kappa_2 G')\frac{\partial h}{\partial x} + (\kappa_2 F' - f')\frac{\partial h}{\partial y}\right]\hat{Z}. \quad (4.1.33b)$$

And from these expressions for $\bar{U}_1$ and $\bar{U}_2$ one can compute the corresponding unit vectors without further dificulties.

Using Eqs. (4.1.22) - (4.1.33) we can then compute the parameters R_1, R_2, $\hat{U}_1$ and $\hat{U}_2$, and hence fully characterize the subreflecton surface in the reflection point neighborhood.

Let's now return to the computation of the field reflected at the point P, shown in Fig. 4.2. The reflection of a spherical wave, emanating from S, in the subreflecton surface element of Fig. 4.4 is depicted in Fig 4.5 through a few relevant rays. The vectors $\bar{S}_i$ and $\bar{S}_R$ are given by

$$\bar{S}_i = (x - x_s)\hat{X} + (Y - Y_s)\hat{Y} + (z - z_s)\hat{Z}, \quad (4.1.34a)$$

$$\bar{S}_R = (X_p - X)\hat{X} + (Y_p - Y)\hat{Y} + (z_p - z)\hat{Z}, \quad (4.1.34b)$$

And correspond to the incident and reflected rays at the point X, Y, and Z.

After the incident spherical wavefront

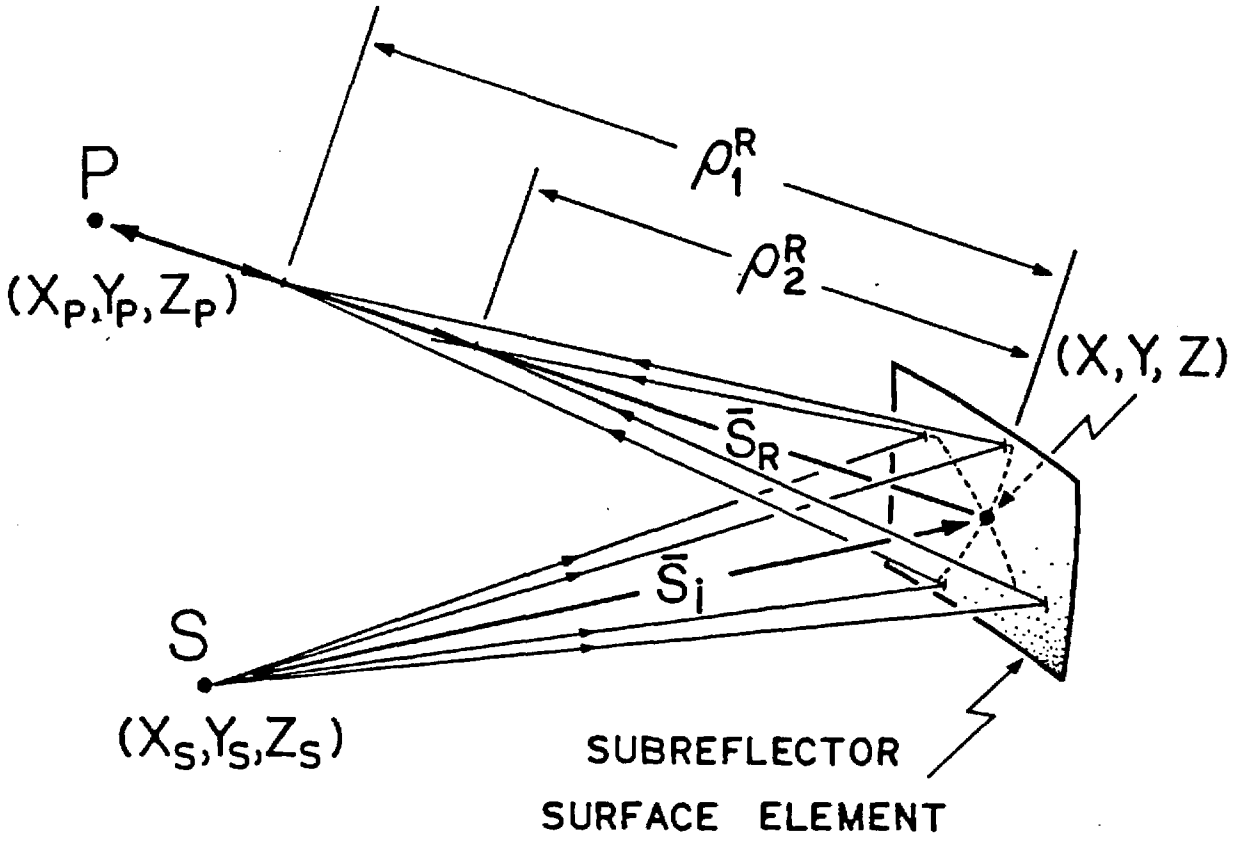


Fig. 4.5 - Reflection of a spherical wave by the subreflector.

reflects on the subreflector a new wavefront shape is generated. This reflected wavefront is, in general, not spherical, but since its surface is analytical, it is then fully described by two principal radii of curvature and two principal directions. Of importance here are only the principal radii of curvatures, shown as ρ_1^R and ρ_2^R in Fig. 4.5. They are given by the expression [1.3]

$$\frac{1}{\rho_{1,2}^R} = \frac{1}{s_i} + \frac{1}{\cos\theta_i}\left[\frac{\sin^2\theta_2}{R_1} + \frac{\sin^2\theta_1}{R_2}\right] \pm$$

$$\pm\sqrt{\frac{1}{\cos^2\theta_i}\left[\frac{\sin^2\theta_2}{R_1} + \frac{\sin^2\theta_1}{R_2}\right] - \frac{4}{R_1 R_2}}\,, \qquad (4.1.35)$$

where

$$\cos\theta_i = -\hat{S}_i \cdot \hat{N}_S\,, \qquad (4.1.36a)$$

$$\sin^2\theta_1 = 1 - (\hat{S}_i \cdot \hat{U}_1)^2\,, \qquad (4.1.36b)$$

$$\sin^2\theta_2 = 1 - (\hat{S}_i \cdot \hat{U}_2)^2\,. \qquad (4.1.36c)$$

Once the principal radii of curvature of the reflected wavefront are known the electric and magnetic fields at P, $\bar{E}_R$ and $\bar{H}_R$ respectively, can be obtained using the ray propagation formulas [1.2]

$$\bar{E}_R = \bar{E}_R\Big|_{S_R=0}\sqrt{\frac{1}{1+S_R/\rho_1}}\sqrt{\frac{1}{1+S_R/\rho_2}}\,e^{-j2\pi S_R}\,, \qquad (4.1.37a)$$

$$\bar{H}_R = \frac{1}{\eta}\,\hat{s}_R \times \bar{E}_R \,. \qquad (4.1.37b)$$

Where $\bar{E}_R|_{s_R=0}$ is the electric field immediately after reflection, given by Snell's law [1.2]:

$$\bar{E}_R|_{s_R=0} = -\bar{E}_E + 2(\hat{N}_S \cdot \bar{E}_E)\hat{N}_S \,. \qquad (4.1.38)$$

In the above $\bar{E}_E$ is the electric field produced by the feed element, located at S, at the reflection point [given by Eq. (3.10a)].

Equation (4.1.37) adopts the convention that ρ_1 and/or ρ_2 are negative if a corresponding caustic exist along the reflected ray path (ρ_1 and ρ_2 shown in Fig. 4.5 are then both negative). The correct ρ_1 and ρ_2 signs are given automatically by Eq. (4.1.35), provided that R_1 and R_2 are taken negative if the corresponding centers of curvature fall in the surface side where the reflection is taken place — the signs of R_1 and R_2 are also given automaticaly, by Eqs. (4.1.22).

As a final remark, the $+\pi/2$ rd phase shifts due to caustic crossings (the Gouy phase shifts [4.2]) are introduced automaticaly by Eq. (4.1.37), provided that it is used exactly in the form shown — the square roots must not be combined or moved to the denominator.

4.1.4- Diffracted-Field Evaluation

In this section we evaluate the subreflecton scattering contribution due to an arbitrary diffraction point. We consider once more the geometry of Fig. 4.1, with the subreflector rim described in the parametric form introduced in Sec. 4.1.2. Also, the diffraction point location is assumed known, and its corresponding parametric coordinate t given.

Consider Fig. 4.6 showing a section of the subreflector rim containing our diffraction point with coordinates X, Y, and Z [determined by substituting t in Eqs. (4.1.13)]. As a first step towards the evaluation of the diffracted field we recall that, in the vicinity of a point, an arbitrary curve can be approximated by an osculating circle [4.4]. We then approximate the subreflector rim, about the diffraction point, by a circle. This circle is shown in Fig. 4.6 and its corresponding parametes can be determined by evaluating the following equations at the diffraction point [4.4], [4.5].

$$R_E = \frac{|\bar{\rho}_s'|^3}{|\bar{\rho}_s' \times \bar{\rho}_s''|} , \qquad (4.1.39a)$$

$$\hat{N}_E = \hat{b} \times \hat{e} , \qquad (4.1.39b)$$

where

$$\hat{e} = -\frac{\bar{\rho}_s'}{|\bar{\rho}_s'|} ,$$

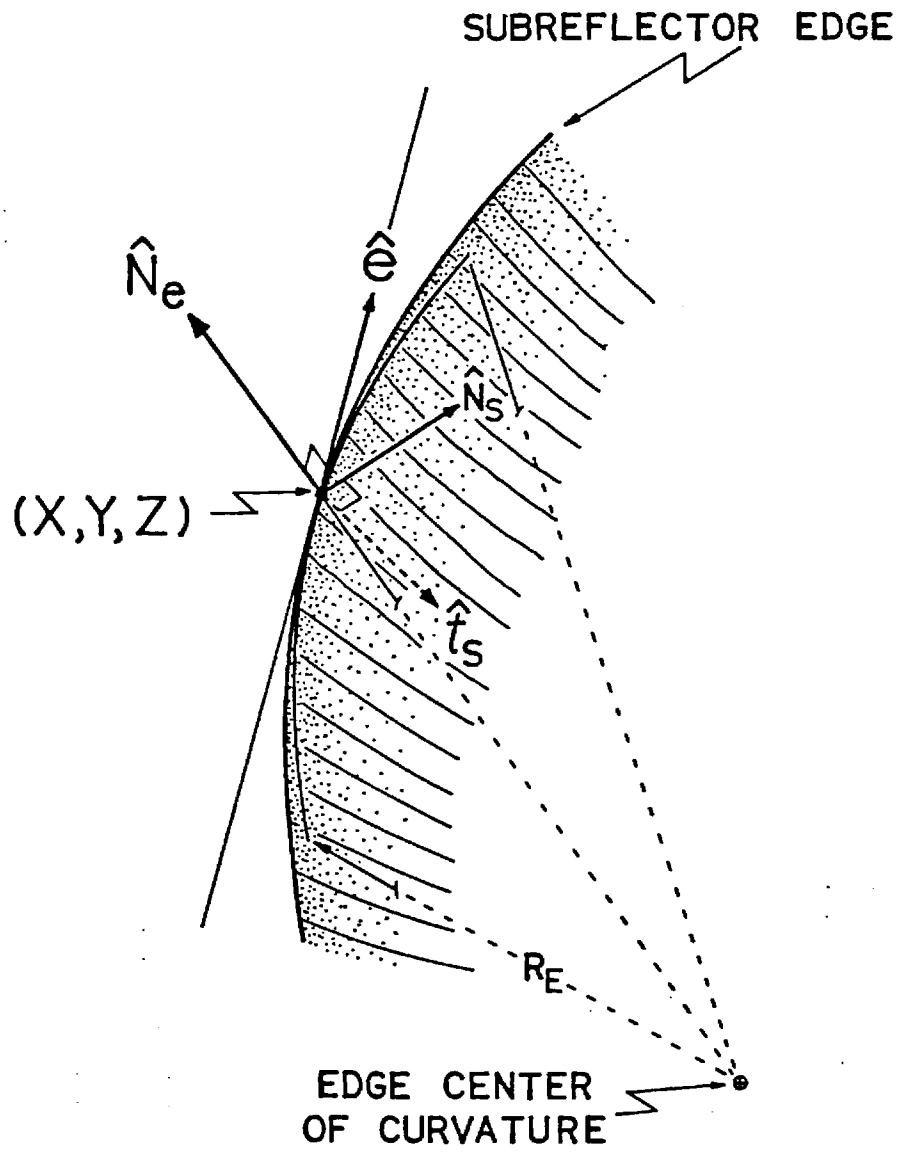


Fig. 4.6 – Approximation of the subreflector rim, about the diffraction point, by a circle.

$$\hat{b} = \frac{\bar{R}_s' \times \bar{R}_s''}{|\bar{R}_s' \times \bar{R}_s''|} , \qquad (4.1.40a)$$

$$\bar{R}_s' = \frac{d\bar{R}_s}{dt} , \qquad (4.1.40b)$$

$$\bar{R}_s'' = \frac{d^2\bar{R}_s}{dt^2} , \qquad (4.1.40c)$$

AND $\bar{R}$ IS A VECTOR FROM THE PRINCIPAL COORDINATE SYSTEM ORIGEN TO A SUBREFLECTOR RIM POINT, OBTAINED USING

$$\bar{R}_s = X(t)\hat{X} + Y(t)\hat{Y} + Z(t)\hat{Z} \qquad (4.1.41)$$

WITH $X(t)$, $Y(t)$ AND $Z(t)$ GIVEN BY EQS. (4.1.13).

IN EQS. (4.1.39) R_E IS THE RADIUS OF THE OSCULATING CIRCLE AND $\hat{N}_E$ IS AN UNIT VECTOR POINTING AWAY FROM THE CIRCLE CENTER — SINCE THE CIRCLE TOUCHES THE DIFFRACTION POINT, THESE TWO PARAMETERS ARE SUFFICIENT FOR ITS CHARACTERIZATION.

TWO OTHER VECTORS ARE ALSO SHOWN IN FIG. 4.6, THE SURFACE UNIT NORMAL $\hat{N}_S$, AND THE UNIT VECTOR $\hat{t}_s$ TANGENT TO THE SURFACE,

$$\hat{t}_s = \hat{e} \times \hat{N}_S . \qquad (4.1.42)$$

THESE TWO VECTORS ARE ALSO DEFINED AT THE DIFFRACTION POINT AND CHARACTERIZE THE SUBREFLECTOR SURFACE — THEY WILL BE USEFUL THROUGHOUT THIS SECTION.

SINCE DIFFRACTION IS A LOCALIZED PHENOMENO THE DIFFRACTED FIELDS PRODUCED BY BOTH THE SUBREFLECTOR AND OSCULATING CIRCLE RIMS CAN BE CONSIDERED IDENTICAL FOR ALL

practical purposes. Our problem can then be reduced to a simpler one, namely the computation of the field diffracted by a point with parametric coordinate t located on a circular rim characterized by the parameters $R_E, \hat{N}_E, \hat{e}$, and $\hat{N}_S$. Note that we are assuming in here that the rim is an edge with negligible thickness – an excellent approximation for most subreflector rims encountered in practice.

The evaluation of the field diffracted by a circular wedge is presented with sufficient detail in [1.3]. Due to this fact we give below only the specific results relevant to our subreflector edge diffraction geometry.

Consider Fig 4.7 showing the same subreflector section of Fig. 4.6. Depicted there is a ray emanating from S and being diffracted in a ray going through the field point P. These two rays are described by the vectors $\bar{S}_i$ and $\bar{S}_D$, where

$$\bar{S}_i = (X(t)-X_s)\hat{X} + (Y(t)-Y_s)\hat{Y} + (Z(t)-Z_s)\hat{Z}, \quad (4.1.43a)$$

$$\bar{S}_D = (X_p-X(t))\hat{X} + (Y_p-Y(t))\hat{Y} + (Z_p-Z(t))\hat{Z}. \quad (4.1.43b)$$

The computation of the diffracted field is greately simplified if the incident and diffracted fields are described in two ray-fixed coordinate systems defined by the unit vectors

$$\hat{\phi}_o' = \frac{\hat{e}\times\hat{S}_i}{|\hat{e}\times\hat{S}_i|}, \quad (4.1.44a)$$

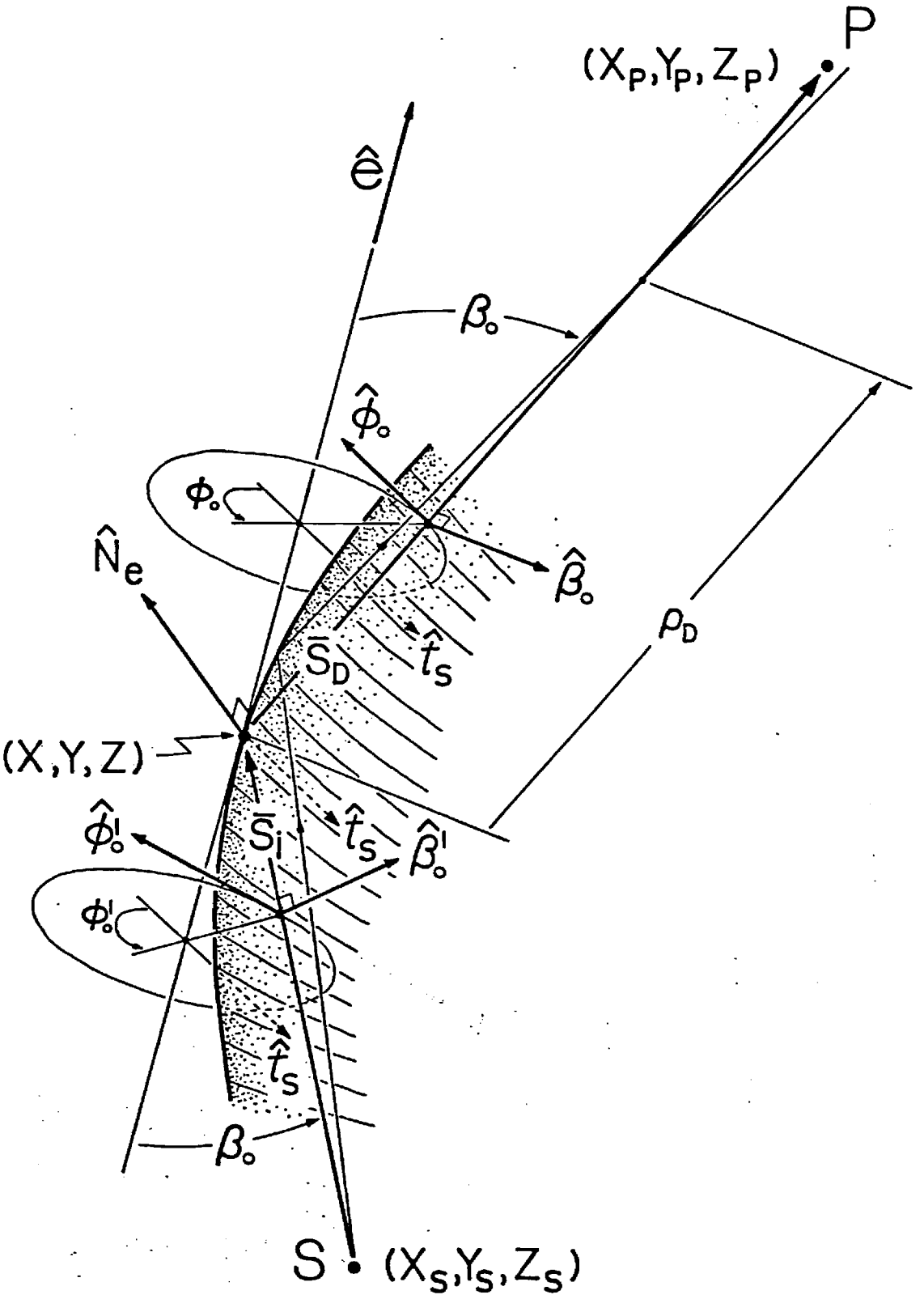


Fig. 4.7 - Diffraction of a spherical wave by the subreflecton edge.

$$\hat{\beta}_0' = \hat{s}_i \times \hat{\phi}_0' \,, \tag{4.1.44b}$$

$$\hat{\phi}_0 = \frac{\hat{s}_D \times \hat{e}}{|\hat{s}_D \times \hat{e}|} \,, \tag{4.1.44c}$$

$$\hat{\beta}_0 = \hat{s}_D \times \hat{\phi}_0 \,. \tag{4.1.44d}$$

These vectors are shown in Fig. 4.7 attached to the incident and diffracted rays. Note that although the angles ϕ_0' and ϕ_0 are in general different, the angles β_0 and β_0' are identical – a consequence of the generalized Fermat's principle.

Using the ray fixed coordinate system, the electric and magnetic field diffracted at P can be evaluated using [1.3]

$$\bar{E}_D = E^D_{\phi_0}\left[(\hat{\phi}_0\cdot\hat{x})\hat{x} + (\hat{\phi}_0\cdot\hat{y})\hat{y} + (\hat{\phi}_0\cdot\hat{z})\hat{z}\right] +$$

$$+ E^D_{\beta_0}\left[(\hat{\beta}_0\cdot\hat{x})\hat{x} + (\hat{\beta}_0\cdot\hat{y})\hat{y} + (\hat{\beta}_0\cdot\hat{z})\hat{z}\right], \tag{4.1.45a}$$

$$\bar{H}_D = \frac{1}{\eta}\,\hat{s}_D \times \bar{E}_D \,, \tag{4.1.45b}$$

where

$$\begin{matrix} E^D_{\phi_0} \\ E^D_{\beta_0} \end{matrix} = -\frac{1}{\sqrt{2\pi}} \begin{matrix} E^i_{\phi_0'}\left[\sqrt{K}\,D_H\right] \\ E^i_{\beta_0'}\left[\sqrt{K}\,D_S\right] \end{matrix} \sqrt{\frac{1}{s_D(1+s_D/\rho_D)}}\; e^{-j2\pi s_D} \,, \tag{4.1.46}$$

AND

$$E^i_{\phi_0'} = \bar{E}_E \cdot \hat{\phi}_0' \ , \qquad (4.1.47a)$$

$$E^i_{\beta_0'} = \bar{E}_E \cdot \hat{\beta}_0' \ . \qquad (4.1.47b)$$

In the above formulas $\bar{E}_E$ is the electric field produced at the diffraction point by the spherical wave source located at S. $\bar{E}_E$ is then a single array element radiated electric field, given by Eq. (3.10a).

In Eq. (4.1.46) ρ_0 is the diffracted ray caustic distance, shown in Fig 4.7, and given by

$$\frac{1}{\rho_0} = \frac{1}{s_i} - \frac{\hat{N}_e \cdot (\hat{s}_i - \hat{s}_D)}{R_E \sin^2\beta_0} \ , \qquad (4.1.48)$$

where

$$\sin^2\beta_0 = 1 - (\hat{s}_D \cdot \hat{e})^2 \ . \qquad (4.1.49)$$

And $\sqrt{K}\, D_H$ and $\sqrt{K}\, D_S$ are the hard and soft diffraction coefficients, respectively computed using

$$\sqrt{K}\, D_{H,S} = \frac{-e^{-j\pi/4}}{2\sqrt{2\pi}\sin\beta_0} \left\{ \frac{F[2KL^i \cos^2(\frac{\phi_0 - \phi_0'}{2})]}{\cos(\frac{\phi_0 - \phi_0'}{2})} \pm \right.$$

$$\left. \pm \frac{F[2KL^R \cos^2(\frac{\phi_0 + \phi_0'}{2})]}{\cos(\frac{\phi_0 + \phi_0'}{2})} \right\} \qquad (4.1.50)$$

In order to obtain $\sqrt{K}\, D_H$ and $\sqrt{K}\, D_S$ using Eq. (4.1.50) a few results are still needed. They are given below.

KL^i AND KL^R ARE distance PARAMETERS GIVEN by

$$KL^i = \frac{2\pi S_D \sin^2\beta_0}{(1+S_D/S_i)}, \qquad (4.1.51a)$$

$$KL^R = \frac{2\pi S_D (1+S_D/\rho_e^R)\sin^2\beta_0}{(1+S_D/\rho_1^R)(1+S_D/\rho_2^R)}, \qquad (4.1.51b)$$

where the CAUSTIC distance ρ_e^R, OF the DIFFRACTED RAY IN the direction OF the RAY that REFLECTS AT the DIFFRACTION POINT, IS EVALUATED USING

$$\frac{1}{\rho_e^R} = \frac{1}{S_i} - \frac{2(\hat{N}_S\cdot\hat{N}_e)(\hat{S}_i\cdot\hat{N}_S)}{R_E\sin^2\beta_0}, \qquad (4.1.52)$$

AND the PRINCIPAL RADII OF CURVATURE OF the WAVEFRONT REFLECTED AT the DIFFRACTION POINT, ρ_1^R AND ρ_2^R, ARE computed USING Eq. (4.1.35).

The ANGLES ϕ_0' AND ϕ_0 ARE CALCULATED, IN A 0 to 2π rd RANGE, USING either

$$\phi_0' = \arccos(\hat{N}_S\cdot\hat{\phi}_0'), \qquad (4.1.53a)$$

$$\phi_0 = \arccos(\hat{N}_S\cdot\hat{\phi}_0), \qquad (4.1.53b)$$

OR

$$\phi_0' = 2\pi - \arccos(\hat{N}_S\cdot\hat{\phi}_0'), \qquad (4.1.54a)$$

$$\phi_0 = 2\pi - \arccos(\hat{N}_S\cdot\hat{\phi}_0). \qquad (4.1.54b)$$

Which ONE IS USED depend ON the VALUE OF the test PARAMETER

$$\tau' = \hat{t}_s \cdot \hat{\phi}_0' . \qquad (4.1.55a)$$

If T' is positive ϕ_0' is evaluated using Eq. (4.1.53a), and if it is negative Eq. (4.1.54'a) is used instead. Similarly for ϕ_0 but using the test parameter

$$\tau = \hat{t}_s \cdot \hat{\phi}_0 . \qquad (4.1.55b)$$

The function $F[2KL\cos^2\sigma]$ apearing in Eq. (4.1.50) is the so called transition function a name derived from the fact that it is responsible for a smooth trinsition of the total GTD field as the shadow and reflection boundaries are crossed. In the above F argument L has been used in place of L^i or L^R, and σ is either $(\phi_0 - \phi_0')/2$ or $(\phi_0 + \phi_0')/2$. The transition function is given by, for positive KL values,

$$\left.\frac{F[2KL\cos^2\sigma]}{\cos\sigma}\right|_{KL>0} = 2j\sqrt{2KL}\, e^{j2KL\cos^2\sigma} \times$$

$$\times f_s(KL,\sigma) \cdot \mathrm{Sign}(\cos\sigma) , \qquad (4.1.56)$$

where $f_s(KL,\sigma)$ is the Fresnel integral

$$f_s(KL,\sigma) = \int_{\sqrt{2KL}}^{\infty} e^{-j\zeta^2} d\zeta . \qquad (4.1.57)$$

And for negative KL values by [4.6]

$$\left.\frac{F[2KL\cos^2\sigma]}{\cos\sigma}\right|_{KL<0} = \frac{|F[|2KL\cos^2\sigma|]|}{\cos\sigma} \times$$

$$\times \exp\left\{-j\left[\text{Phase of } F[|2KL\cos^2\sigma|]\right\}\right. \qquad (4.1.58)$$

For $U = 2KL\cos^2\sigma$ laying in the range $0.01 \leq U < 10$ we have computed the required transition function values using Eq. (4.1.56) and the Fresnel integral approximations given in [4.7]. Outside this range we have used the approximations [1.3]

$$\frac{F[U]}{\cos\sigma} \simeq \text{sign}(\cos\sigma)\sqrt{2KL}\left[\sqrt{\pi}-2\sqrt{U}\,e^{j\pi/4}(1-jU/3)\right]e^{j(\pi/4+U)} \qquad (4.1.59)$$

For $U \leq 0.01$, and

$$\frac{F[U]}{\cos\sigma} \simeq \frac{1}{\cos\sigma}\left\{1+\frac{1}{U}\left[\frac{j}{2}+\frac{1}{U}\left[-\frac{3}{4}+\frac{1}{U}\left(-j\frac{15}{8}+\frac{1}{U}\frac{75}{16}\right)\right]\right]\right\} \qquad (4.1.60)$$

For $U \geq 10$. Eqs. (4.1.59) and (4.1.60) have error less than 10^{-7} and 10^{-4}, respectively.

In this section we are adopting the same convention used in Sec. 4.1.3 for the caustic distances. According to it ρ_0, ρ_e^R, ρ_1^R and ρ_2^R are negative if a corresponding caustic is crossed by the ray. This sign convention is automatically provide by the formulas yielding the caustic distances. Also, in the same maner as Eq. (4.1.37a), Eq. (4.1.46) automaticaly

includes a $+\pi/2$ phase shift if the diffracted ray crosses a caustic, provided that the square root is not moved to the denominator of the term involving S_D.

At this point we have the complete procedure for computing the subreflector scattering according to the Uniform Geometrical Theory of Diffraction. The total electric and magnetic fields, $\bar{E}_s'$ and $\bar{H}_s'$ respectively, produced at a point P by a subreflector illuminated by a single feed array element can then be determined using

$$\bar{E}_s'(P) = \sum_{\substack{\text{REFLECTION} \\ \text{POINTS}}} \bar{E}_R(P) + \sum_{\substack{\text{DIFFRACTION} \\ \text{POINTS}}} \bar{E}_D(P), \quad (4.1.61a)$$

$$\bar{H}_s'(P) = \sum_{\substack{\text{REFLECTION} \\ \text{POINTS}}} \bar{H}_R(P) + \sum_{\substack{\text{DIFFRACTION} \\ \text{POINTS}}} \bar{H}_D(P), \quad (4.1.61b)$$

where $\bar{E}_R(P)$ and $\bar{H}_R(P)$ are given by Eqs. (4.1.37) and $\bar{E}_D(P)$ and $\bar{H}_D(P)$ by Eqs. (4.1.45). And by summing up the $\bar{E}_s'(P)$ and $\bar{H}_s'(P)$ of each feed element the final subreflector scatted fields $\bar{E}_s(P)$ and $\bar{H}_p(P)$ are obtained in mathematical form

$$\bar{E}_s(P) = \sum_{\substack{\text{ARRAY} \\ \text{elements}}} \bar{E}_s'(P), \quad (4.1.62a)$$

$$\bar{H}_s(P) = \sum_{\substack{\text{ARRAY} \\ \text{elements}}} \bar{H}_s'(P). \quad (4.1.62b)$$

4.2 - Physical Optics

As previously mentioned another technique for computing the subreflector scattering is the so called Physical Optics method. In this technique an approximation is used for the current density induced over the subreflector surface by the complete feed array, and then this current is integrated over the free-space tensor Green's function to yield the subreflector scattered field. In this section we present this other approach.

According to [1.2] the electric and magnetic fields $\bar{E}(P)$ and $\bar{H}(P)$, respectively, produced at a given point P by a surface current distribution $\bar{J}_S$, distributed over a surface S, are given by

$$\bar{E}(P) = \frac{1}{4\pi} \iint_S \left[-j\omega\mu_0 \psi \bar{J}_S + \frac{1}{j\omega\epsilon_0} (\bar{J}_S \cdot \nabla) \nabla \psi \right] ds, \quad (4.1.63a)$$

$$\bar{H}(P) = \frac{1}{4\pi} \iint_S \left[\bar{J}_S \times \nabla \psi \right] ds, \quad (4.1.63b)$$

where

$$\psi = \frac{e^{-jkR}}{R}, \quad (4.1.64a)$$

$$\nabla \psi = \left(jk + \frac{1}{R} \right) \psi \hat{R}, \quad (4.1.64b)$$

$$(\bar{J}_s\cdot\nabla)\nabla\psi = \left\{\left[\frac{3}{R}\left(jK+\frac{1}{R}\right)-K^2\right](\bar{J}_s\cdot\hat{n})\hat{n} - \frac{\bar{J}_s}{R}\left(jK+\frac{1}{R}\right)\right\}\psi, \quad (4.1.64c)$$

where $\bar{n}$ is a vector from $\bar{J}_s$ to P. ω is the angular frequency of the current $\bar{J}_s$. K is the free space phase constant. And ϵ_0 and μ_0 are the free space permittivity and permeability respectively.

In the form show the dimensions of Eqs. (4.1.63) and (4.1.64) are not in wavelengths. Since in this work all dimensions are in wavelengths it is convenient to transform Eqs. (4.1.63) and (4.1.64) into the following forms:

$$\bar{E}(P) = -\frac{\partial\mu}{4\pi}\iint_S\left\{2\pi(\lambda\psi)\bar{J}_s + \frac{1}{2\pi}[\lambda^3(\bar{J}_s\cdot\nabla)\nabla\psi]\right\}ds, \quad (4.1.65a)$$

$$\bar{H}(P) = \frac{1}{4\pi}\iint_S[\bar{J}_s\times(\lambda^2\nabla\psi)]\,ds, \quad (4.1.65b)$$

where

$$\lambda\psi = \frac{e^{-j2\pi n}}{n}, \quad (4.1.66a)$$

$$\lambda^2\nabla\psi = \left(j2\pi+\frac{1}{n}\right)(\lambda\psi)\hat{n}, \quad (4.1.66b)$$

$$\lambda^3(\bar{J}_s\cdot\nabla)\nabla\psi = \left\{\left[\frac{3}{n}\left(j2\pi+\frac{1}{n}\right)-(2\pi)^2\right](\bar{J}_s\cdot\hat{n})\hat{n} - \frac{\bar{J}_s}{n}\left(j2\pi+\frac{1}{n}\right)\right\}(\lambda\psi). \quad (4.1.66c)$$

Now all dimensions are in wavelengths. η is the free-space wave impedance.

Application of Eqs. (4.1.65) to evaluate the subreflecton scattered fields is done according to Fig. 4.8. Note that all coordinate systems shown in it have been previously discussed.

Subjected to the physical-optics approximation, the $\bar{J}_S$ induced at an arbitrary point on the subreflecton surface (with coordinates ρ_F, θ_F, ϕ_F) by the feed array is

$$\bar{J}_S = 2\hat{N}_S \times \bar{H}_A \qquad (4.1.67)$$

Where $\hat{N}_S$ is the unit normal to the subreflecton surface at the $\bar{J}_S$ point [given by Eq. (2.13)]. And $\bar{H}_A$ is the the total magnetic field produced by the feed array at the $\bar{J}_S$ point (from Sec. 3).

Recalling that in the ρ_F, θ_F, ϕ_F spherical coordinate system the subreflecton has its rim at $\theta_F = \theta_E$, it is convenient to integrate Eqs. (4.1.65) in the variables θ_F and ϕ_F (the integration limits will be constants). The area element on the subreflecton surface, in the ρ_F, θ_F, ϕ_F system is:

$$dS_s = \frac{\rho_F^2 \sin\theta_F \, d\phi_F \, d\theta_F}{|\hat{N}_S \cdot \hat{\rho}_F|} \qquad (4.1.68)$$

And the electric and magnetic fields, $\bar{E}_s$ and $\bar{H}_s$ respectively, scattered by the subreflecton at the point P can be obtained by numerically integrating Eqs. (4.1.65), rewritten bellow for reference

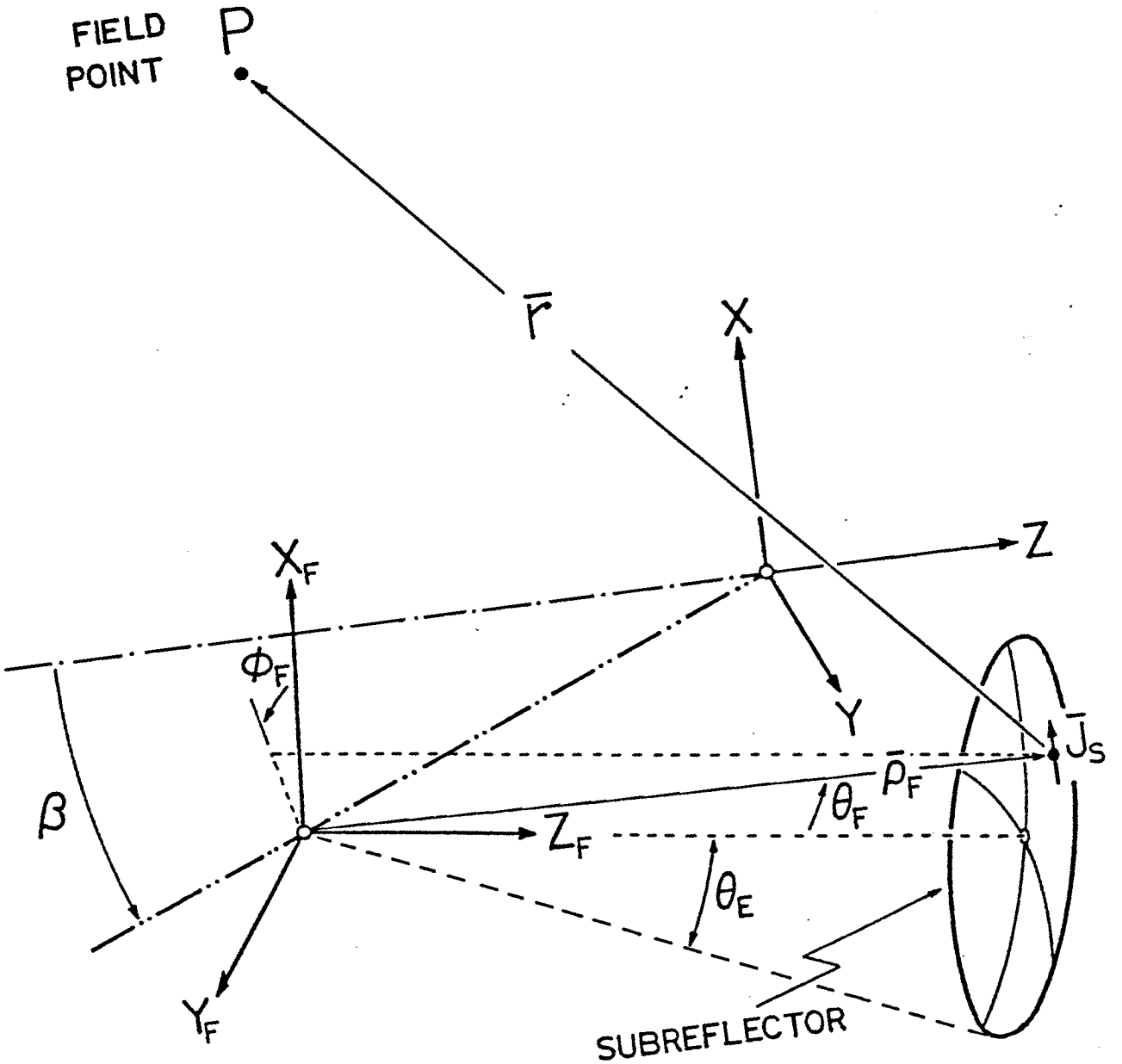


Fig. 4.8 - Geometry for computation of the subreflector scattering using Physical Optics.

$$\overline{E}_s(P) = \frac{-\jmath\eta}{4\pi} \int_0^{\theta_E} \int_0^{2\pi} \left\{ 2\pi(\lambda\psi)\overline{J}_s + \frac{1}{2\pi}\left[\lambda^3(\overline{J}_s\cdot\nabla)\nabla\psi\right] \right\} \frac{\rho_F^2 \sin\theta_F}{|\hat{N}_s\cdot\hat{\rho}_F|} \, d\phi_F \, d\theta_F \qquad (4.1.69a)$$

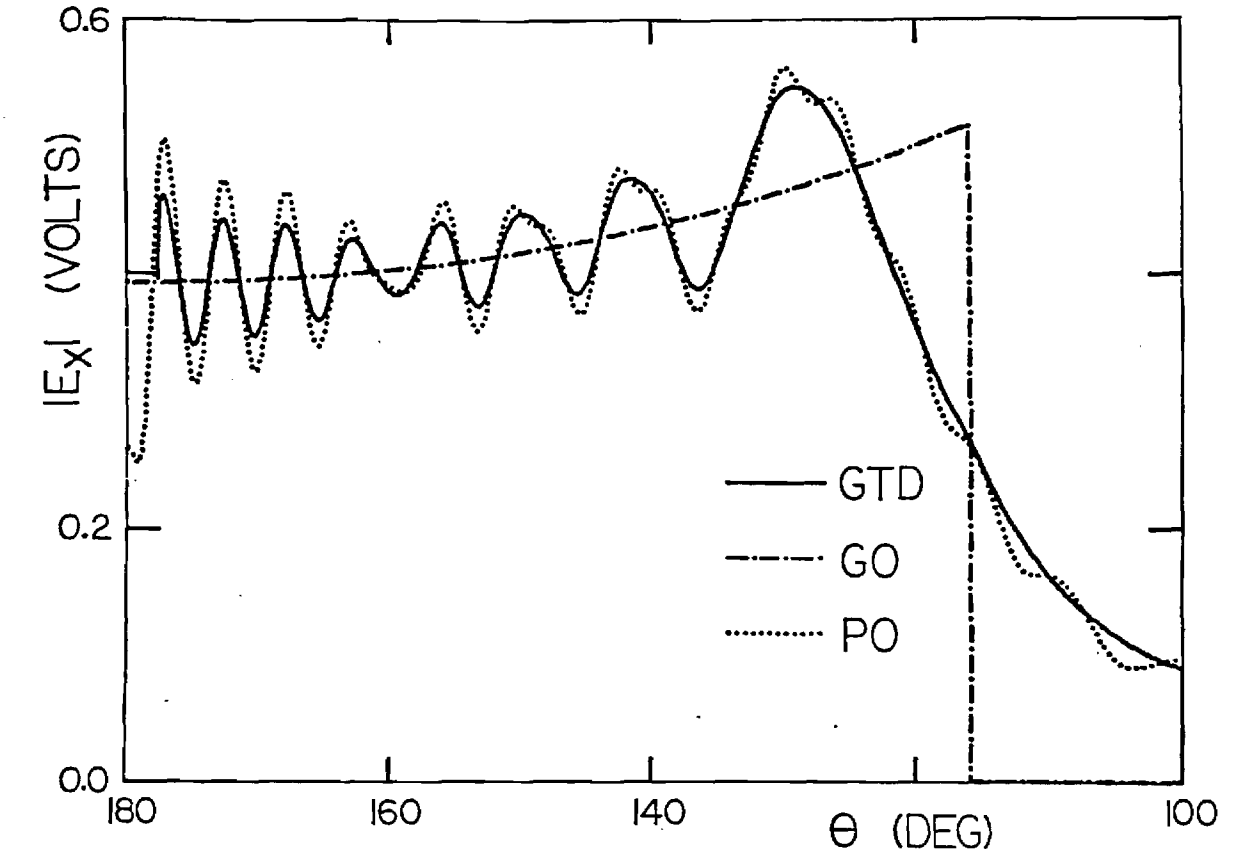
$$\overline{H}_s(P) = \frac{1}{4\pi} \int_0^{\theta_E} \int_0^{2\pi} \left[\overline{J}_s \times (\lambda^2 \nabla\psi)\right] \frac{\rho_F^2 \sin\theta_F}{|\hat{N}_s\cdot\hat{\rho}_F|} \, d\phi_F \, d\theta_F \qquad (4.1.69b)$$

4.3 - Typical Results

In this section we present a sample of results obtained using the methods discussed in Secs. 4.1 and 4.2. Our purpose in here is both to illustrate the level of agreement that can be expected between the GTD and PO computations, and to validate our numerical procedures.

All results shown below correspond to the scattering from a hyperboloidal surface with eccentricity $e = 2.295200$, inter-focal distance $2c = 30.00000\,\lambda$, and depression angle $\ell = 0$. The illumination is provided by a single raised-cosine-type feed (see Sec. 3), pointed in the $\hat{z}$ direction ($A_z = E_L = 0$), polarized along the $\hat{x}$ direction ($\phi_{E_0} = 0$), without pattern taper (exponents $E = H = 0$), and excited with EEXCIT = 1. The computations were carried out with code SORADC (presented in Sec. 6), modified to evaluate the subreflector scattered electric field over an sphere of radius R. This sphere is centered at the principal coordinate system origen, and all field values are given without the corresponding e^{-jkR}/R spherical wave factor.

The first set of results is shown in Fig. 4.9, together with the subreflector geometry — a 25λ diameter axially symmetric section of the hyperboloid has been selected by choosing $\alpha = 0$ and $\theta_E = 27.60392°$. The feed is located at the hyperboloid focus with coordinates $X = Y = 0$ and $Z = -30\lambda$. As seen, the GTD and PO computations agree very closely, both



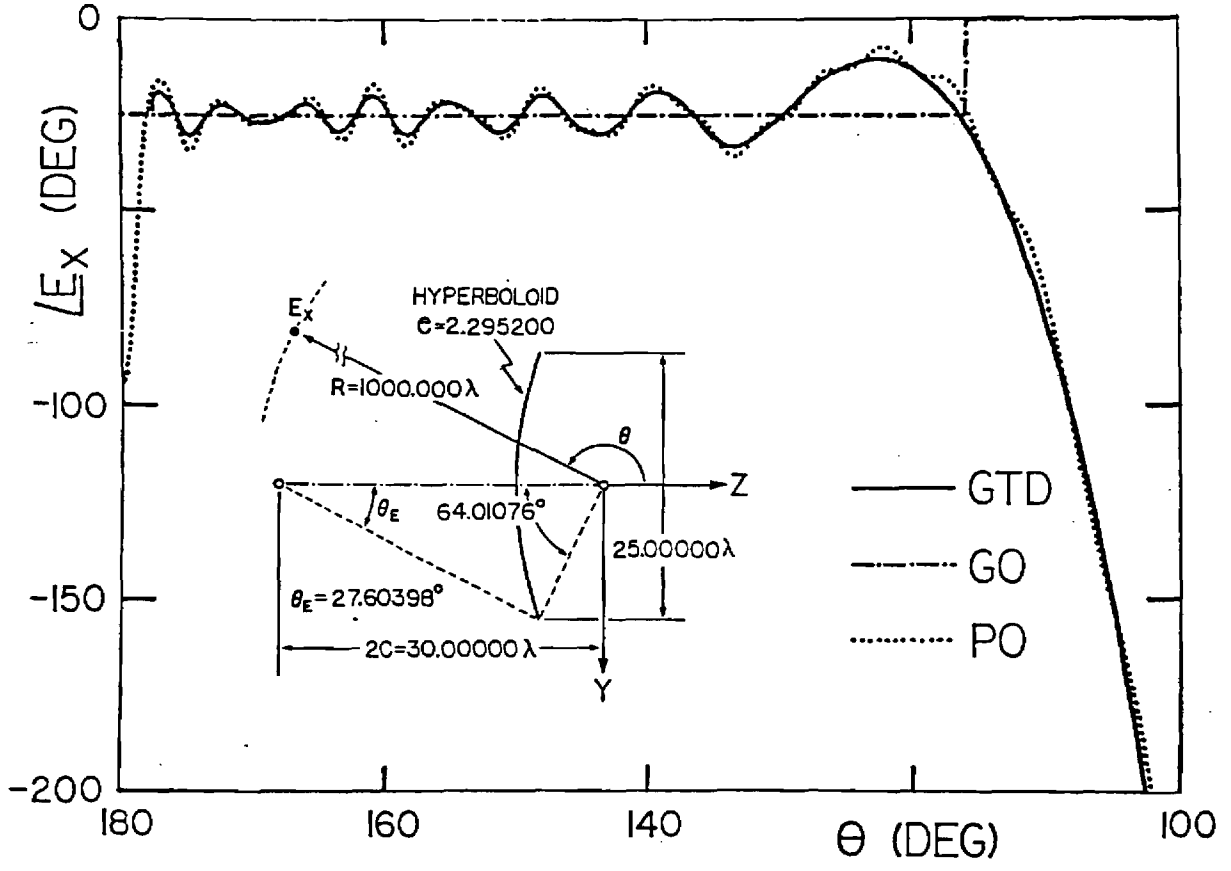
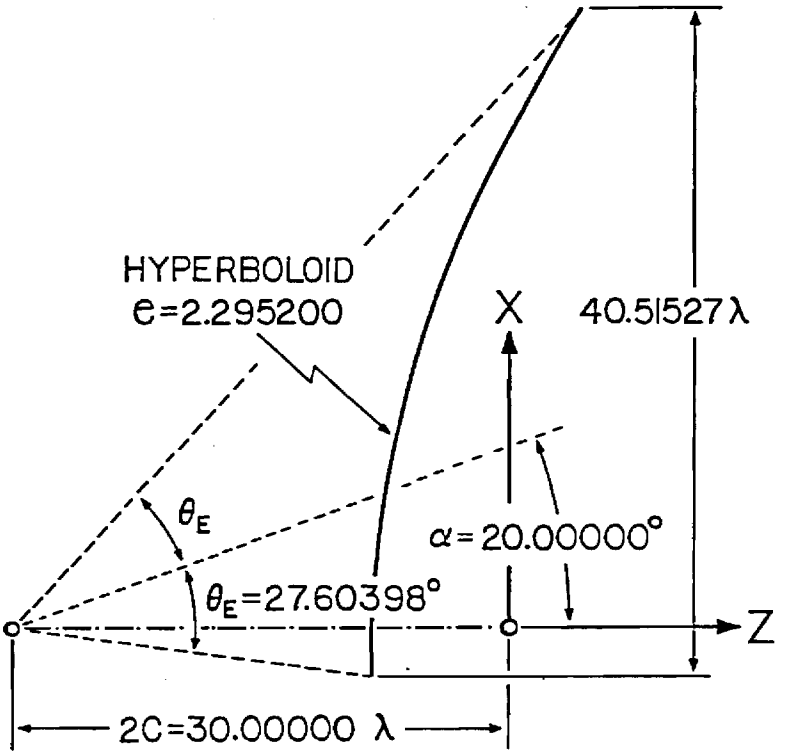


Fig. 4.9 - Axially Symmetric Subreflector Scattered Electric Field.

in magnitude and phase, and the computation using only reflected rays (curved labeled GO, for geometrical optics) present an useful approximation for the scattered field within the range $\theta > 116°$. The GTD curve is not shown for θ in the vicinity of 180° due to the fact that, in this region, the rim yields no well defined stationary phase points – the z axis is a caustic for the diffracted rays and hence GTD breaks down. Totally independent scattering results for this geometry are available in the literature [1.4], [4.8], and our computations agree very well with them.

A more general test case is presented in Figs. 4.10 and 4.11. It corresponds to an offset section of the already described hyperboloid, specified by $X = 20°$ and $\theta_E = 27.60398°$. The feed is scanned and located at the point with $X = 4$, $Y = -2$ and $Z = -30$, so that the geometry has no symmetry. Independent results for the scattering produced by this subreflector is also available in [4.8], and they show an excellent agreement with ours (note that the feed used in [4.8] has EEXCIT = -120π and the e^{-jkR}/R factor was not factored out in the scattering results given there).

The small discontinuities shown in Fig. 4.10, in the vicinity of $\Omega = \pm 77°$, are caused by diffracted rays emanating, from the diffraction point, almost tangent to the subreflector surface. The GTD formulation used is not valid when this occurs [1.3]. However, this is not a limitation for the present work,



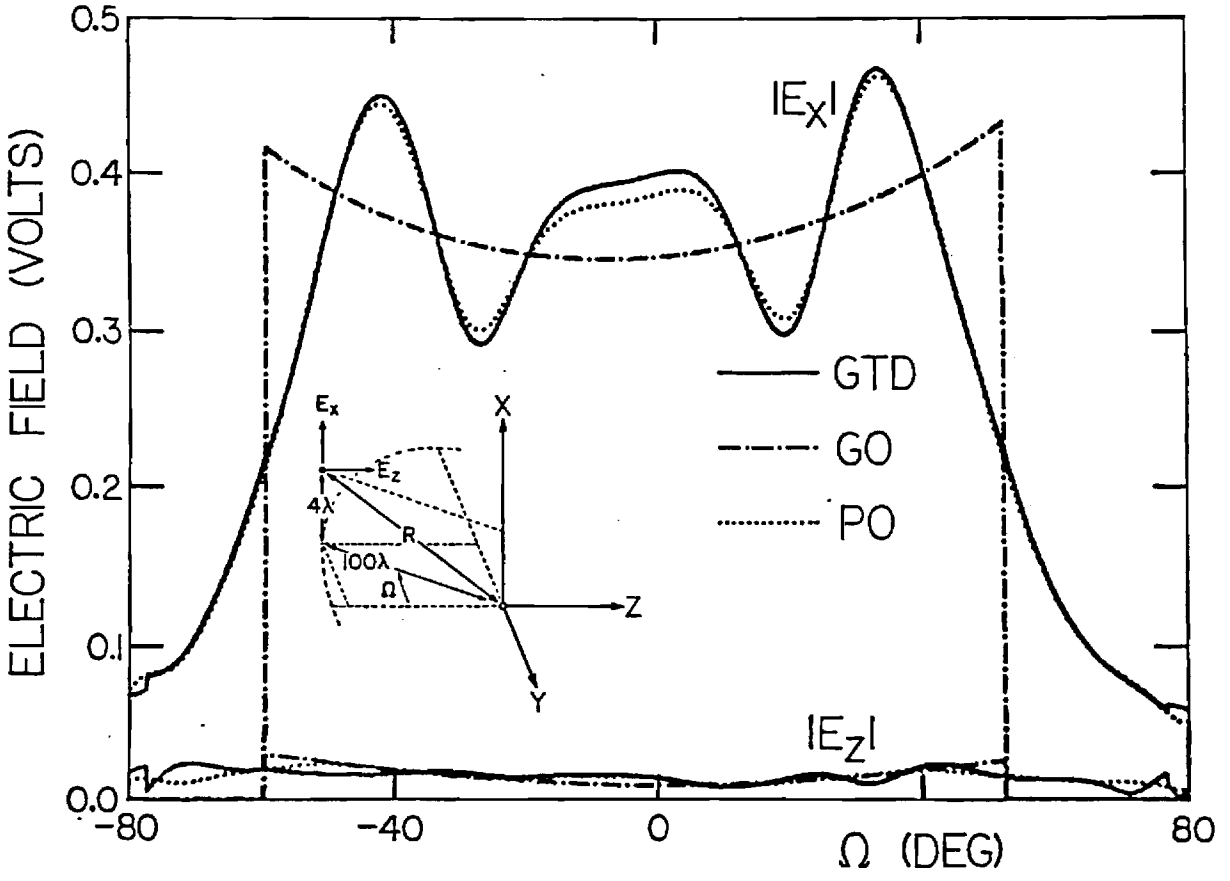
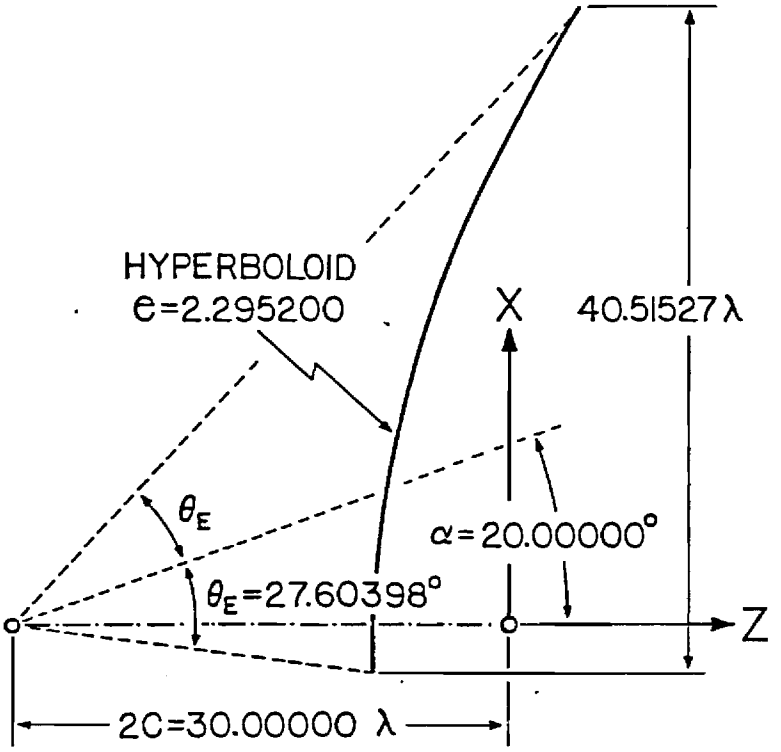


Fig. 4.10 - Offset Hyperboloidal Subreflector Scattered Electric Field. Plane with X=4λ



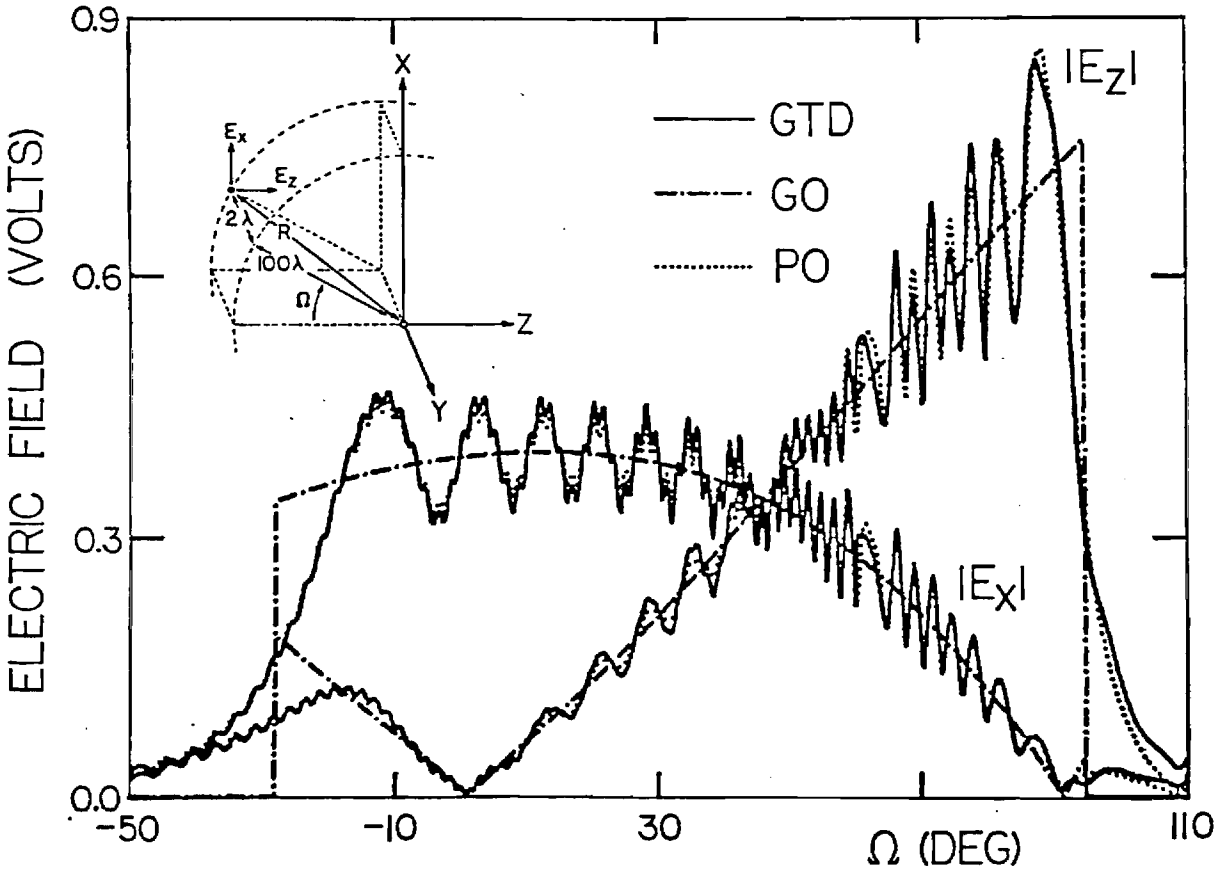


Fig. 4.11 - Offset Hyperboloidal Subreflector Scattered Electric Field. Plane with X=-2λ.

since in practical antennas rays of this type are seldom faced when computing the main reflector illumination of practical antennas.

The advantage of computing the subreflector scattering using GTD, as opposed to PO (at least in this single feed application), can be appreciated by noting that the GTD curves of Fig. 4.9 were generated ~160 times faster than the corresponding PO ones [each curve has 321 points, and 80 integration steps were used in the radial angular coordinate when computing Eq. (4.1.69)]. Also, the GO curve was evaluated 660 times faster than PO. For the curves shown in Figs. 4.10 and 4.11 the speed difference was even greater; GTD and GO were 2000 and 12000 times faster than PO, respectively. This impressive speed difference between PO and the other two methods is mainly due to the offset subreflector used, which in addition of being much larger than the one of Fig. 4.9, also has a receeding top edge — two factors that demand a high number of integration steps for evaluating Eq. (4.1.69) with acceptable accuracy (300 integration steps were used in the radial angular coordinate and each curve shown in Figs. 4.10 and 4.11 has 641 points).

5- Main-reflector Scattering

As mentioned in the introduction, over the pattern region of interest in this work, the field radiated by the dual-reflector antenna system is mainly generated by the main-reflector scattering contribution. In this section we discuss the evaluation of the main-reflector scattered field using the physical optics technique.

In Sec. 4.2 the equations for computing the fields radiated by a surface current distribution were presented, namely Eqs. (4.1.65). Here the field point P is in the far-zone region of the antenna, in this circunstance Eqs. (4.1.65) reduce to:

$$\bar{E}(P) = \frac{-j\eta}{2}\left[\frac{e^{-j2\pi R}}{R}\right]\iint_S \left[\bar{J}_s - (\bar{J}_s\cdot\hat{R})\hat{R}\right] e^{j2\pi \bar{\rho}_M\cdot\hat{R}}\, ds \qquad (5.1a)$$

$$\bar{H}(P) = \frac{1}{\eta}\,\hat{R}\times\bar{E}(P) \qquad (5.1b)$$

Where S is again the surface over which the current density $\bar{J}_s$ rests. $\bar{R}$ and $\bar{\rho}_M$ are vectors from the principal coordinate system origen to the far-field point and to the $\bar{J}_s$ point, respectively.

Figure 5.1 shows the application of Eqs. (5.1) to the main reflector of our antenna system. With the exception of the R, θ, φ spherical system, all the parameter and coordinate systems shown there have been discussed previously.

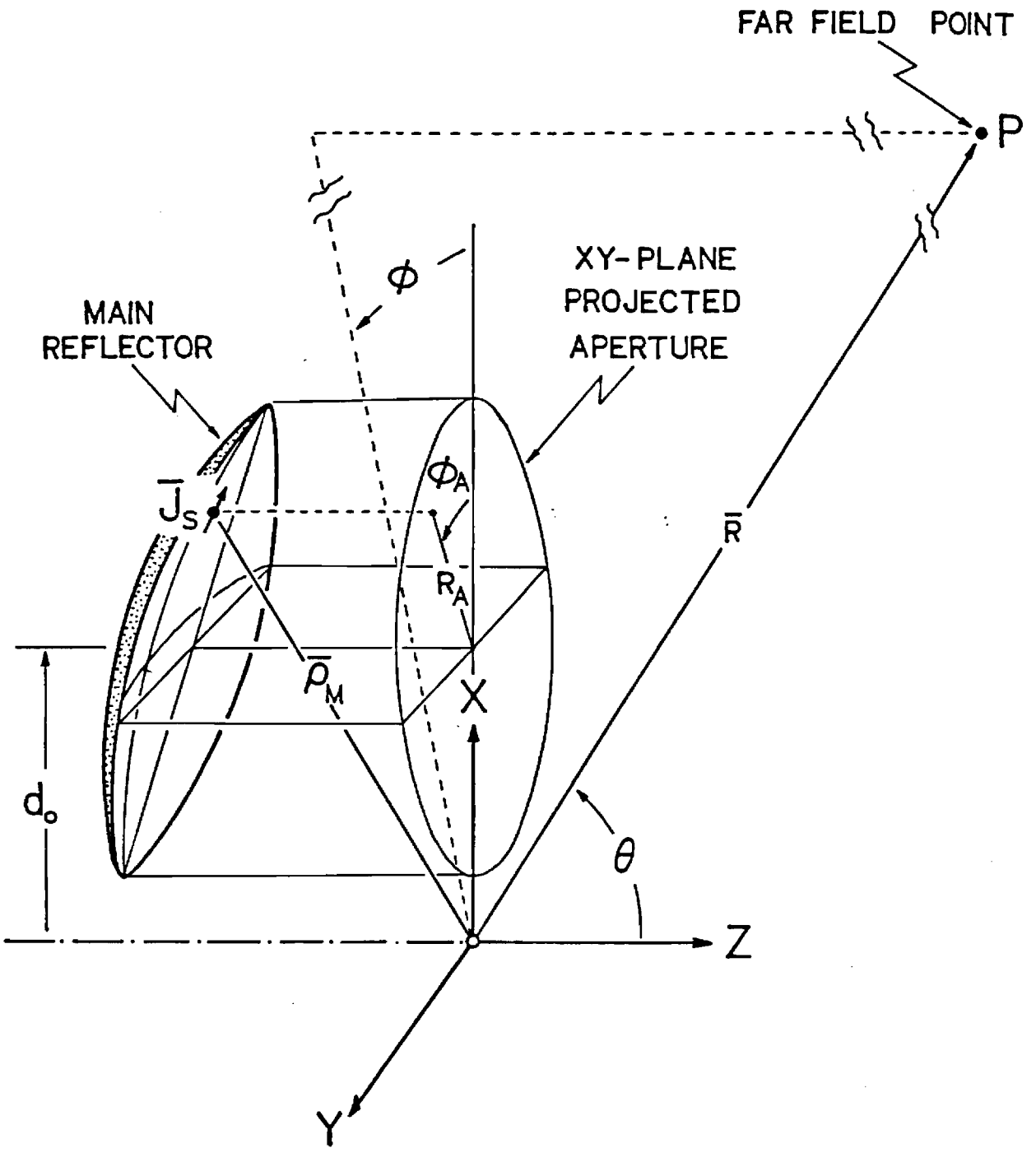


Fig 5.1 - Geometry for computation of the main-reflector far-field scattering.

Using once more the physical optics approximation the surface current density $\bar{J}_S$ induced at the main-reflector point, with coordinates R_A and ϕ_A, by the subreflector scattered field is

$$\bar{J}_S = 2\hat{N}_R \times \bar{H}_S . \qquad (5.2)$$

Where $\hat{N}_R$ is the main-reflector surface unit normal at the $\bar{J}_S$ point [given by Eq. (2.10)], and $\bar{H}_S$ is the magnetic field scattered by the subreflector, also at the $\bar{J}_S$ point. $\bar{H}_S$ is given by either Eq. (4.1.62b) or (4.1.69b), depending on the method selected to compute the subreflector scattering.

The integration of Eqs. (5.1) is more conveniently carried out in the Z_A, R_A, ϕ_A system; due to the fact that the main reflector projected aperture is circular (the integration limits will be constants). The main reflector area element in the R_A, ϕ_A system is:

$$dS_R = \frac{R_A \, d\phi_A \, dR_A}{|\hat{N}_R \cdot \hat{z}|} . \qquad (5.3)$$

With this result the far-field produced by the main-reflector can be written as [from Eqs. (5.1)],

$$\bar{E}_T(P) = -\frac{jM}{2} \int_0^{D/2} \int_0^{2\pi} [\bar{J}_S - (\bar{J}_S \cdot \hat{R})\hat{R}] e^{j2\pi \bar{\rho}_M \cdot \hat{R}} \frac{R_A}{|\hat{N}_R \cdot \hat{z}|} \, d\phi_A \, dR_A , \qquad (5.4a)$$

$$\bar{H}_T(P) = \frac{1}{\eta} \hat{R} \times \bar{E}_T(P) . \qquad (5.4b)$$

Since the bracket term in Eq. (5.1a) is simply the phase and atenuation factor of a spherical wave emanating from the principal system origen, it has been droped in Eq. (5.4a).

In this work Eqs. (5.4) are integrated in the principal system cartesian components, and then the results are transformed into the R, θ, ϕ spherical components of the system of Fig 5.1. This is standard in this type of scattering computations since the radiated fields are TEM in the spherical system. Due to its convenience for practical applications, it is also desirable to have the antenna radiated fields in terms of co-polarized and cross-polarized components according to the widely approved Ludwigs third definition, given in [3.1] as

$$E_{co} = E_\theta \cos(\phi - \phi_{PR}) - E_\phi \sin(\phi - \phi_{PR}), \quad (5.5a)$$

$$E_{cx} = E_\theta \sin(\phi - \phi_{PR}) + E_\phi \cos(\phi - \phi_{PR}), \quad (5.5b)$$

where ϕ_{PR} defines the electric field direction considered to be co-polarized. As an example, if $\phi_{PR} = 0$, on the z-axis the co-polarized component is in the x-axis direction; if $\phi_{PR} = 90°$ it is along the y-axis, and so on. And E_θ and E_ϕ are the θ and ϕ components of $\bar{E}_T$, respectively.

When computing the radiation pattern of directive antennas, the standard procedure is to hold ϕ constant while θ is varied to obtain a pattern cut about the $\theta = \phi = 0°$ direction. The θ and ϕ system is then adequate as long as the direction of interest is the $\theta = \phi = 0°$ direction. When the direction of interest is not aligned with the z-axis (such as in scanned beam applications), the θ, ϕ system becomes inadequate for computing the pattern cuts. In this case a more convenient system is the shifted spherical θ_s, ϕ_s one, shown in Figs. 5.2 and 5.3. This system allows the central direction for the pattern cuts to be specified in any arbitrary direction $\theta = \theta_0, \phi = \phi_0$, and then, by holding $\tilde{\phi}_s$ constant and varying θ_s, the pattern cuts can be obtained.

In Fig. 5.2 the auxiliary coordinate systems required to establish the shifted system are shown. From it the following relations are obtained:

$$X = X'_s \cos\tilde{\phi}_0 - Y'_s \sin\tilde{\phi}_0 \qquad (5.6a)$$

$$Y = X'_s \sin\tilde{\phi}_0 + Y'_s \cos\tilde{\phi}_0 \qquad (5.6b)$$

$$Z = Z'_s \qquad (5.6c)$$

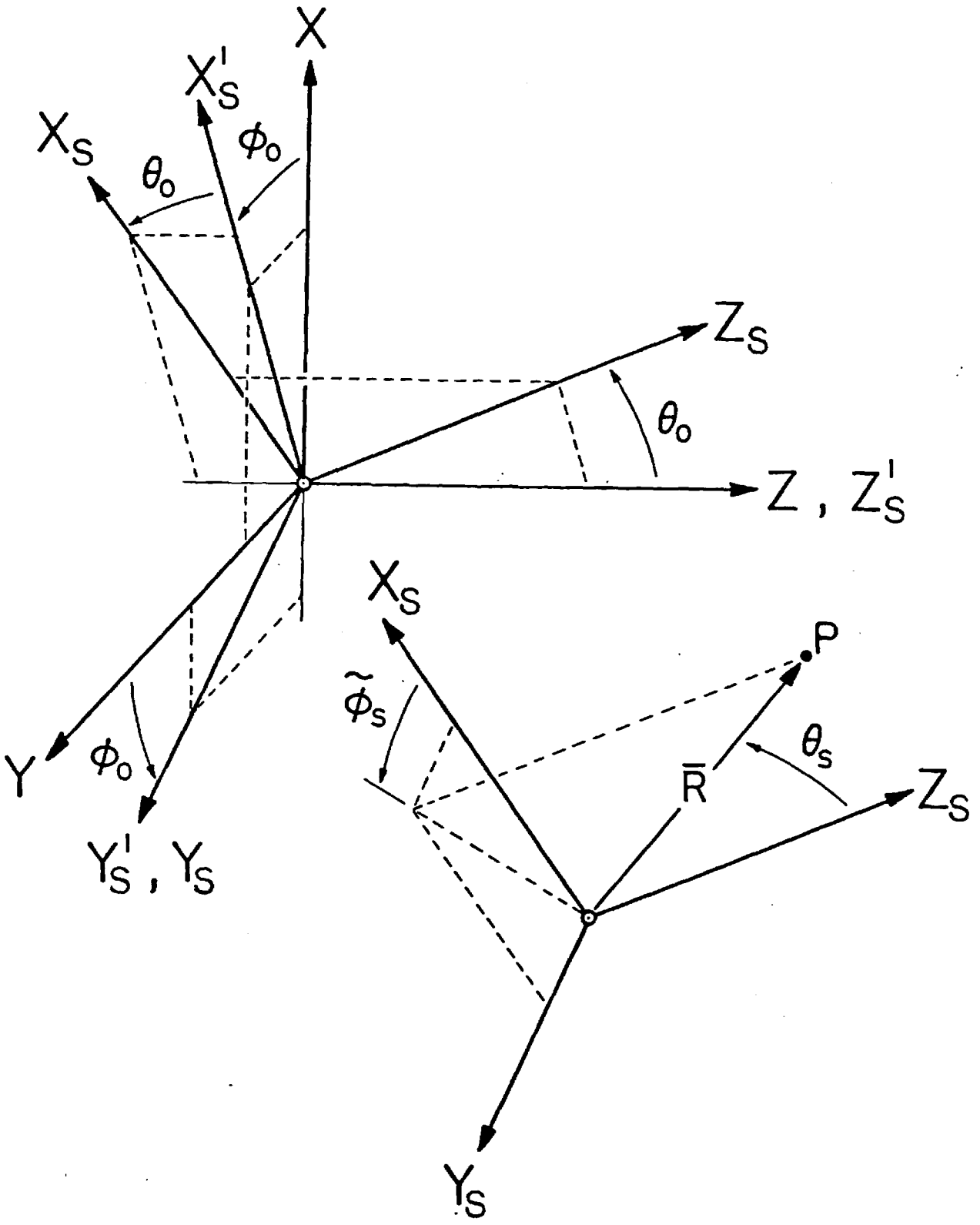


Fig. 5.2 - Shifted spherical coordinate system for pattern computation

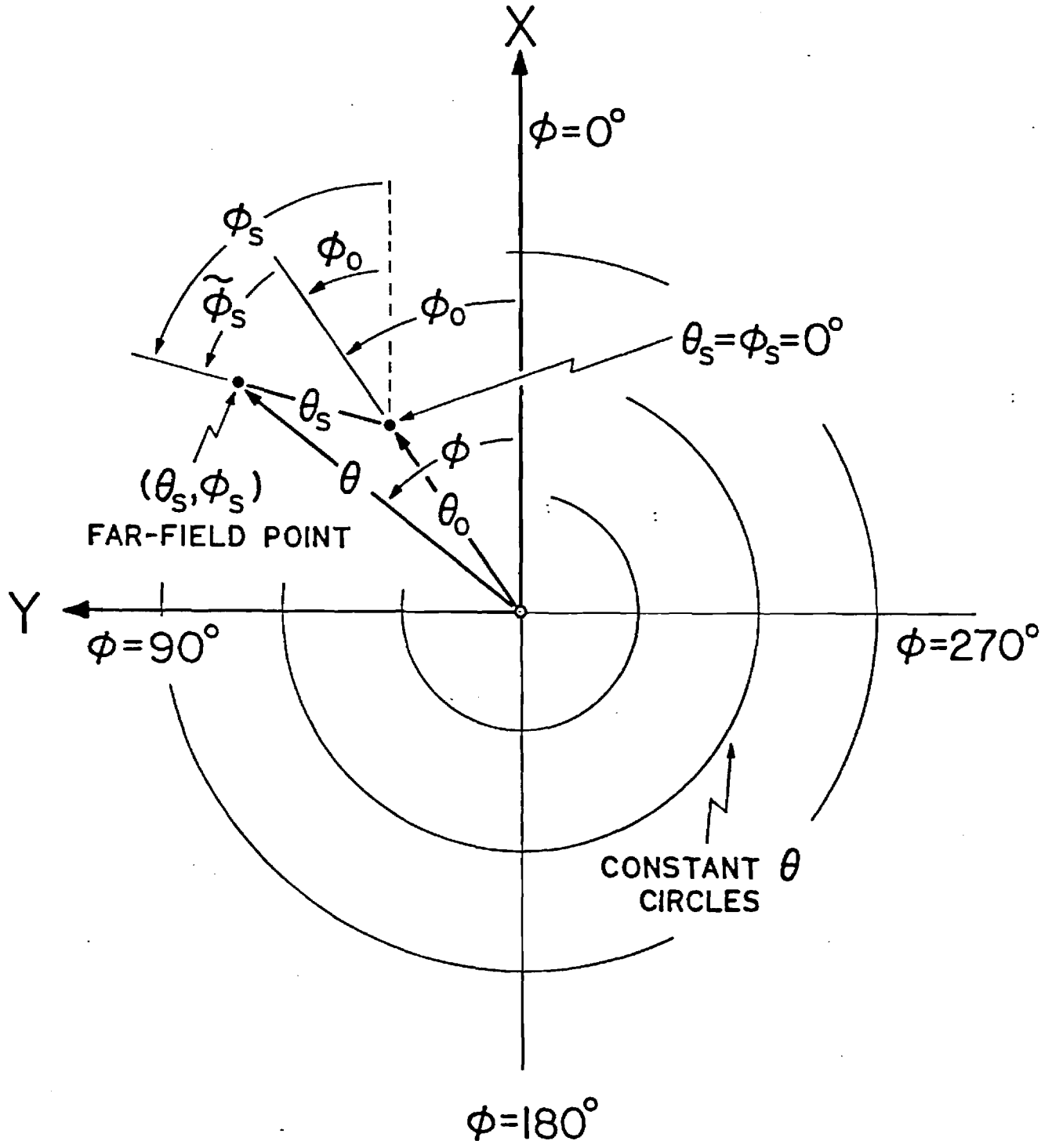


Fig. 5.3 - Shifted spherical coordinate system for pattern computation. (XY-plane projection)

AND

$$X_s' = X_s \cos\theta_0 + Z_s \sin\theta_0 \tag{5.7a}$$

$$Y_s' = Y_s \tag{5.7b}$$

$$Z_s' = -X_s \sin\theta_0 + Z_s \cos\theta_0 \tag{5.7c}$$

Also,

$$X_s = R \sin\theta_s \cos\tilde{\phi}_s \tag{5.8a}$$

$$Y_s = R \sin\theta_s \sin\tilde{\phi}_s \tag{5.8b}$$

$$Z_s = R \cos\theta_s \tag{5.8c}$$

If these three sets of equations are combined with

$$X = R \sin\theta \cos\phi \tag{5.9a}$$

$$Y = R \sin\theta \sin\phi \tag{5.9b}$$

$$Z = R \cos\theta \tag{5.9c}$$

one arrive at the following relations connecting the θ, ϕ and the θ_s, ϕ_s' spherical systems:

$$\theta = \text{ACOS}\left[\cos\theta_s \cos\theta_0 - \sin\theta_s \cos\tilde{\phi}_s \sin\theta_0\right] \tag{5.10a}$$

$$\phi = \text{ATAN}\left[\frac{\sin\theta_s(\cos\tilde{\phi}_s \cos\theta_0 \sin\phi_0 + \sin\tilde{\phi}_s \cos\phi_0) + \cos\theta_s \sin\theta_0 \sin\phi_0}{\sin\theta_s(\cos\tilde{\phi}_s \cos\theta_0 \cos\phi_0 - \sin\tilde{\phi}_s \sin\phi_0) + \cos\theta_s \sin\theta_0 \cos\phi_0}\right] \tag{5.10b}$$

As previously mentioned, the field results computed using Eqs. (5.4) are obtained in the principal cartesian system components, and then are transformed in the θ and ϕ components. If their $\theta_s, \tilde{\phi}_s$

components are desired, the X, Y, Z components can be transformed using the unit vectors:

$$\hat{X} = \hat{X}'_s \cos\phi_0 - \hat{Y}'_s \sin\phi_0 \qquad (5.11a)$$

$$\hat{Y} = \hat{X}'_s \sin\phi_0 + \hat{Y}'_s \cos\phi_0 \qquad (5.11b)$$

$$\hat{Z} = \hat{Z}'_s \qquad (5.11c)$$

and

$$\hat{X}'_s = \hat{X}_s \cos\theta_0 + \hat{Z}_s \sin\theta_0 \qquad (5.12a)$$

$$\hat{Y}'_s = \hat{Y}_s \qquad (5.12b)$$

$$\hat{Z}'_s = -\hat{X}_s \sin\theta_0 + \hat{Z}_s \cos\theta_0 \qquad (5.12c)$$

and

$$\hat{\theta}_s = \hat{X}_s \cos\theta_s \cos\tilde{\phi}_s + \hat{Y}_s \cos\theta_s \sin\tilde{\phi}_s - \hat{Z}_s \sin\theta_s \qquad (5.13a)$$

$$\hat{\tilde{\phi}}_s = -\hat{X}_s \sin\tilde{\phi}_s + \hat{Y}_s \cos\tilde{\phi}_s \qquad (5.13b)$$

Equations (5.6)-(5.13) allow the pattern cuts to be carried on in the $\theta_s, \tilde{\phi}_s$ system, and the field results to be outputed in the θ_s, ϕ_s components. The direction from which $\tilde{\phi}_s$ is measured is shown in Fig. 5.3. Note that this figure is only a projection of the far field sphere in the XY-plane – the field sphere is actually bending away beneath the paper plane, and hence Fig. 5.3 is strictly valid only for arbitrarily small θ-values.

A more convenient azimuth angle for the shifted spherical system is the ϕ_s shown in Fig. 5.3. The relation

between ϕ_s and $\tilde{\phi}_s$ is

$$\phi_s = \tilde{\phi}_s + \phi_0 \quad (5.14)$$

Also, $\hat{\phi}_s = \hat{\tilde{\phi}}_s$ and then the ϕ_s component of the fields is identical to the $\tilde{\phi}_s$ component. From now one in this work what is called the shifted spherical system is always the θ_s, ϕ_s system. The reason why ϕ_s is more convenient than $\tilde{\phi}_s$ is that it is fully compatible with Ludwig's 3rd. polarization definition. This can be best shown through an example. Suppose $\hat{\theta}$ with $\phi = 0$ is the co-polarized direction at the $\theta = 0$ point, then the $\hat{\theta}_s$ direction with $\phi_s = 0$ will also be the co-polarized direction at any point $\theta_s = 0$, for any values of θ_0 and ϕ_0. Fig. 5.4 shows the $\phi_s = 0$ directions over the far-field sphere, note that they are identical to the corresponding equal-polarization directions given by Eqs. (5.5).

As a last result, bellow is given the co-polarized and cross-polarized components of the electric field according to Ludwig's 3rd. definition, in the shifted system:

$$E^s_{co} = E_{\theta_s} \cos(\phi_s - \phi_{PR}) - E_{\phi_s} \sin(\phi_s - \phi_{PR}) \quad (5.15a)$$

$$E^s_{cx} = E_{\theta_s} \sin(\phi_s - \phi_{PR}) + E_{\phi_s} \cos(\phi_s - \phi_{PR}) \quad (5.15b)$$

Note that $E^s_{co} = E_{co}$ and $E^s_{cx} = E_{cx}$ only at the $\theta_s = 0$ point.

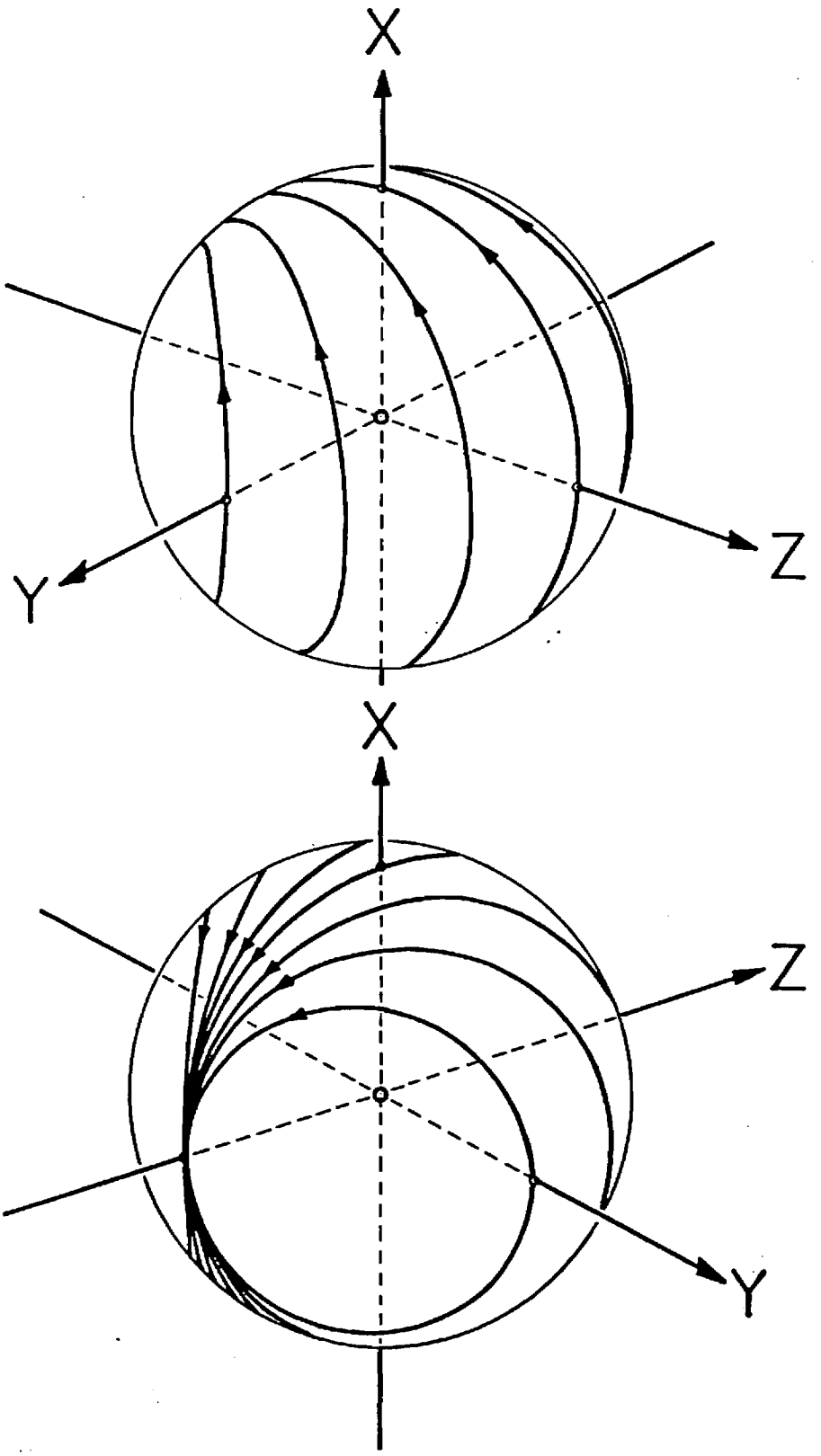


Fig. 5.4 - Shifted spherical coordinate system $\phi_s = 0°$ directions.

6- Computer Program SORADC

The results of the analysis described in the previous sections of this report have been included in the FORTRAN-77 (ANSI X3.9-1978) language computer code SORADC. This code evaluates the scattering characteristics of offset classical dual-reflector antennas. The antenna excitation is provided by an arbitrarily located, oriented, and excited array of linearly polarized raised-cosine-type feeds (described in Sec. 3). Geometrical optics is used to approximate the array fields over the subreflector surface. The subreflector scattered fields are computed using either the Geometrical Theory of Diffraction or the Physical Optics surface current integration methods, according to Secs. 4.1 or 4.2, respectively. The main-reflector scattering is obtained using Physical Optics by evaluating the double integral of the subreflector-induced surface current density. No angles, distances, or unit vectors are approximated. Blocking effects of any kind are not taken into account. Program output includes three arbitrary pattern cuts, gain, and spillover results.

The antenna system far-field radiation pattern is assumed to be generated only by the main-reflector scattering (e.g. subreflector spillover effects are not considered). Also, the main-reflector illumination is assumed provided only by the subreflector scattered field (e.g. direct array feed contribution is ignored).

These assumptions are very reasonable for the antenna systems considered in this work.

All the input parameters required for program SORADC operation are provided by the data file SORADC.DAT. The program output is provided by four files: SORADC.OUT, PACUT1.DAT, PACUT2.DAT, and PACUT3.DAT. File SORADC.OUT is the principal output file and contains all results. PACUT1.DAT, PACUT2.DAT, and PACUT3.DAT are output files intended to be used for radiation pattern plotting, each one contain a table of the co- and cross-polarized radiated fields, as a function of θ_s; one file for each one of the three constant ϕ_s pattern cuts. The shifted spherical system θ_s, ϕ_s used in the cuts is described in Sec. 5. Also, the user has a choice on what spherical system the co- and cross-polarized results are computed, it can be in the θ_s, ϕ_s or in the θ, ϕ (see also Sec. 5).

Program SORADC is almost completely written in single-precision — only the transformations between θ_s, ϕ_s and θ, ϕ, and some of the GTD-related subroutines employ double-precision Arithmetic. Also, all its dimensions are in wavelengths. The majority of the equations used in the code have the same form used when they were presented in this report. In this way the compatibility between theory and code has been guaranteed, and any modifications required in the future can be handled easily.

In the next few sections all the material required to successfully operate program SORADC is presented, and several sample runs are included to help illustrate its usage. The source code is well commented, and additional effort was employed in making the process of following SORADC's operation an easy task. We strogly believe that this goal was acheived, and in case additional information on the code is required, we recomend a closer look at the source code itself.

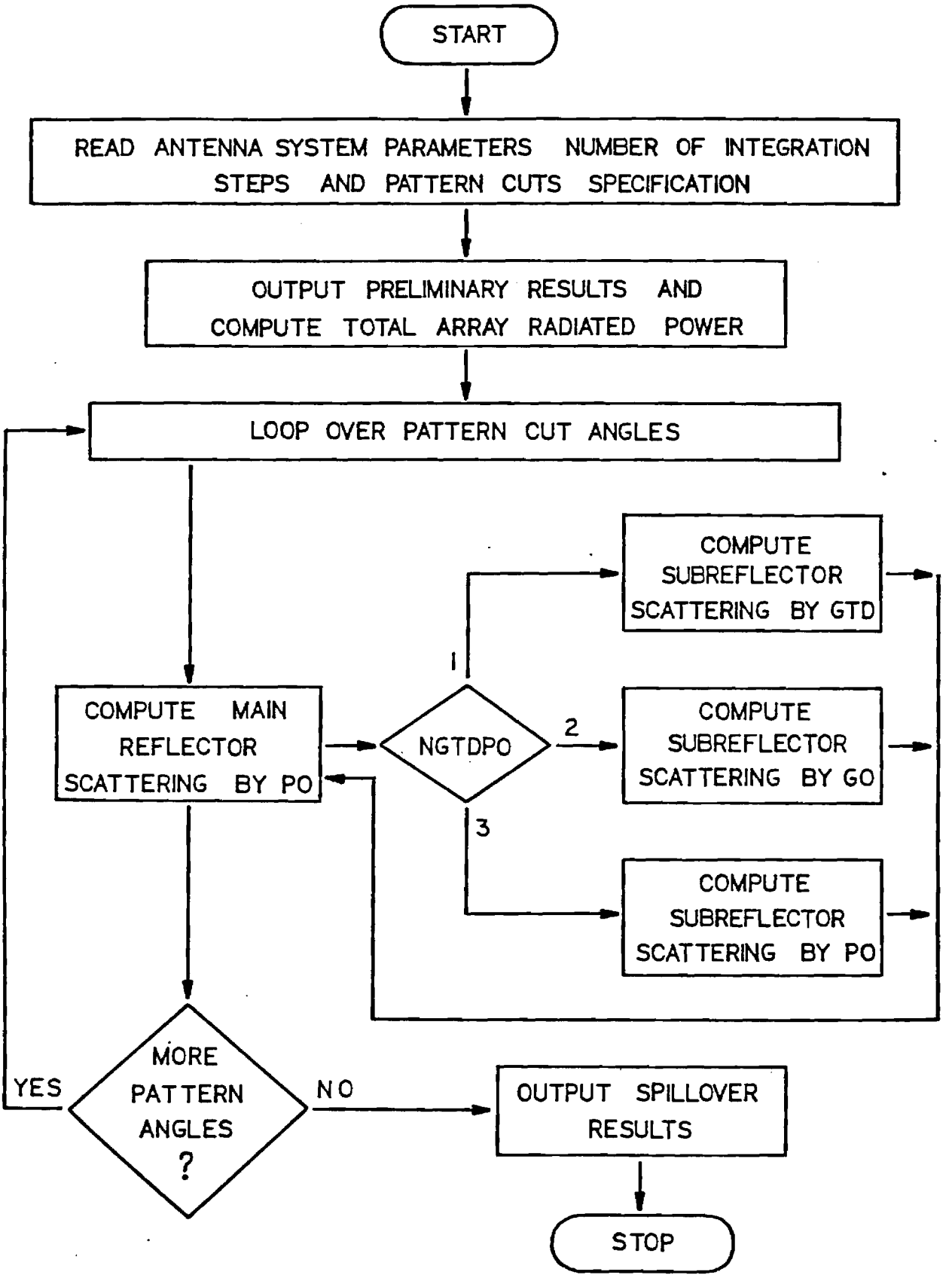


Fig. 6.1- Program SORADC Block Diagram.

6.1 - Operation

In Fig. 6.1 a simplified block diagram of program SORADC operation is given. The program RUN starts by reading the data file SORADC.DAT, which contains the antenna system parameters, the feed array characteristics, information on what method should be employed in computing the subreflector scattering, the number of radial steps used in the numerical integrations, and the desired pattern cuts specifications. SORADC then writes preliminary results in the output file SORADC.OUT, and the computations start. The first result evaluated is the total array radiated power, which is obtained by integrating the array pattern over the far-field sphere. After this SORADC loops over all θ_s values of the three constant-ϕ_s pattern cuts, and evaluates and outputs the corresponding scattered fields. And finally the RUN ends by outputing the antenna spillover results.

As mentioned previously, program SORADC evaluates the subreflector scattering using either the GTD or the PO methods. This choice is controlled by the parameter NGTDPO. If NGTDPO = 1, SORADC determines the scattering using GTD and includes both edge diffracted and surface reflected rays (geometrical optics, or GO, rays). If NGTDPO = 2 only the GO rays are used in computing the scattering. And if NGTDPO = 3 the PO method

is used. The option set by NGTDPO = 2 was included to allow the possibility of using an ideal subreflector whenever desired — in almost all reflector antennas the subreflector rim diffraction is an undesired effect, and this option allows the user to determine its impact. Another reason to have the NGTDPO = 2 option is that, since the diffracted rays are not computed, the program runs significantly faster when it is used, and in many applications still yields usable results for the main-reflector scattering.

When SORADC is operating in the NGTDPO = 1 mode, it sometimes occurs that the main-reflector field point falls too close to a subreflector diffracted ray caustic. In such cases the code is instructed to ignore the edge diffracted rays and to use only the GO rays. Since the GO field represents the dominant contribution to the subreflector scattering (see Sec. 4.3), and the caustic region is usually very small, this procedure is very safe.

The output patterns are always evaluated by holding ϕ_s constant and marching on θ_s with constant step size (see Sec. 5 for more on the θ_s, ϕ_s system). However, the user has choice of the θ_0, ϕ_0 point where all cuts cross, and of the spherical system used to output the field components and gain results. This is controlled by the parameter ISYSTP. If ISYSTP = 1 the field results are the θ and ϕ components (see Fig. 5.1), and the co-

AND CROSS-POLARIZED GAIN RESULTS ARE OBTAINED USING Eqs. (5.5). AND WHEN ISYSTP=2, THE FIELD RESULTS ARE THE θ_s AND ϕ_s COMPONENTS (SEE Fig. 5.3), AND THE CO- AND CROSS-POLARIZED GAIN RESULTS ARE OBTAINED USING Eqs. (5.15)

THREE SURFACE INTEGRALS ARE EVALUATED BY THE PROGRAM: THE TOTAL ARRAY RADIATED POWER INTEGRATION (WITH INTEGRATION VARIABLES θ AND ϕ), THE SUBREFLECTOR SCATTERING INTEGRATION (WITH INTEGRATION VARIABLES θ_F AND ϕ_F), AND THE MAIN-REFLECTOR SCATTERING INTEGRATION (WITH INTEGRATION VARIABLES R_A AND ϕ_A). IN ALL OF THEM THE RADIAL INTEGRAL (THE ONE IN θ, θ_F OR R_A) IS DONE USING SIMPSON'S RULE, AND THE AZIMUTHAL ONE (IN ϕ, ϕ_F OR ϕ_A) USING TRAPEZOIDAL RULE. THE NUMBER OF STEPS USED IN THE RADIAL DIRECTIONS IS SPECIFIED BY THE VARIABLES NTARAY, NRASUB, AND NRDPAR, RESPECTIVELY — ALL THREE SHOULD BE EVEN, BUT IF THEY AREN'T SORADC WILL USE THE SMALLEST CLOSEST EVEN NUMBER. THE NUMBER OF STEPS USED IN THE AZIMUTHAL DIRECTION IS EVALUATED INTERNALLY BY SORADC; THIS IS DONE SO THAT THE INTEGRATION POINTS ARE UNIFORMLY DISTRIBUTED OVER THE INTEGRATION SURFACE. WHEN THE CODE IS USING GTD OR GO TO COMPUTE THE SUBREFLECTOR SCATTERING, THE SUBREFLECTOR INTEGRATION IS STILL EVALUATED WITH NRASUB RADIAL STEPS, BUT ONLY TO DETERMINE THE SUBREFLECTOR REFLECTED POWER (FOR SPILLOVER COMPUTATION PURPOSES).

THE PROBLEM OF SPECIFYING THE VALUES OF NTARAY, NRASUB, AND NRDPAR IS AN IMPORTANT ONE — THESE THREE PARAMETERS

control the program accuracy and speed. The best values for them are clearly the smallest ones that still provide sufficient engineering accuracy. Unfortunately they depend strongly on the antenna and array parameters, and hence no single rule exist for their determination. As a final result for this section, a few formulas will be presented that allow for their estimation. They give good results as long as the antenna is not excessively out of focus and the array feed has a well behaved pattern (a pattern with a well defined beam, like a raised-cosine-type pattern). Under these conditions the required number of integration steps can be estimated by:

$$NTARAY \simeq 5 \times \frac{180}{\text{ARRAY HPBW}} , \qquad (6.1.1a)$$

$$NRASUB \simeq 2 \times D_{SUB} , \qquad (6.1.1b)$$

$$NRDPAR \simeq 5 \times D \times \sin \theta_M . \qquad (6.1.1c)$$

Where Array HPBW is the array half-power-beamwidth (in degrees), D_{SUB} is the largest subreflector projected diameter (the projection plane should be chosen accordingly), and θ_M is the largest value of the θ_s pattern cut angle (measured from the beam peak direction)

Equations (6.1.1b) and (6.1.1c) have been obtained observing the oscillatory behavior of the scattering integrals integrand phase [An extensive discussion of Eq. (6.1.1c) can be found in [6.1]]. From the amplitude behavior another pair of equations can be obtained,

$$\text{NRASUB} = \text{NRDPAR} \simeq \frac{5 \times \Theta_E}{\text{ARRAY HPBW}}, \qquad (6.1.2)$$

with Θ_E in degrees. The values of NRASUB and NRDPAR that should be used is the largest value given by Eqs. (6.1.1b), (6.1.1c) and (6.1.2). Also, if the subreflector scattering is being evaluated using GTD or GO (NGTDPO = 1 or 2), NRASUB is the one given by Eq. (6.1.2).

The above equations are only estimates, and should be used with great caution. For instance, in the computation of Figs. 4.10 and 4.11 (PO curves), the estimate given by Eq. (6.1.1b) was not sufficient due to the particular geometry of the subreflector top edge.

6.2- Data Files

a) Input data file SORADC.DAT

This is the only input data file used by the program SORADC. It contains the scattering computation control variables, the pattern cuts parameters, the antenna system dimensions, and the feed array specifications.
All dimensions used in the data file are given in wavelengths, and all angles are in degrees. Bellow is a detailed description of each one of its lines.

LINE 1 : TITLE1. Any 80-column title for identification purposes.

LINE 2 : TITLE2. Any 80-column title for identification purposes.

LINE 3 : NGTDPO, NRDPAR, NRASUB, NTARAY. Scattering computation control variables. If NGTDPO = 1, the subreflector scattering is computed using GTD. If NGTDPO = 2 only GO is used. And If NGTDPO = 3 PO is used. NRDPAR, NRASUB, and NTARAY are the number of radial integration steps used in the two dimensional integrations. NRDPAR is for the main-reflector, NRASUB for the subreflector, and NTARAY for the array radiated power evaluation — All three must be

even numbers. Note that NRADVB is used even if NGTDPO ≠ 3, since the subreflector reflected power is still obtained by integration in this case.

LINE 4: THETOD, PHIOD, ISYSTP, PHIRFD. Radiation pattern cut parameters. THETOD and PHIOD is the direction about which the pattern cuts are taken, the θ_o and ϕ_o angles of Figs. 5.2 and 5.3. ISYSTP selects in which system the electric field components and the gain results are outputed. ISYSTP = 1 for results in the θ, ϕ system of Fig. 5.1, and ISYSTP = 2 for results in the θ_s, ϕ_s system (see Figs. 5.2 and 5.3). PHIRFD is the co-polar direction of the electric field vector (when $\theta_o = \phi_o = 0$, the $\theta_s = \phi_s = 0$ pattern point co-polarization is X-polarized if PHIRFD = 0, Y-polarized if PHIRFD = 90°, and so on).

LINE 5: PHIS1, PHIS2, PHIS3, THETSI, DTHETS, NTHETS. Radiation pattern cut parameters. PHIS1, PHIS2, and PHIS3 are the constant ϕ_s angles of the three pattern cuts. THETSI is the starting θ_s angle of the cuts, DTHETS is the θ_s angular increment, and NTHETS the number of θ_s values in each cut.

LINE 6: IFEED, E, H. Array feed elements specification. IFEED defines the feed type and should be set

equal to 1. E and H are the raised cosine E and H plane pattern exponents, respectively [see Eqs. (3.9)].

LINE 7: D, F, DØ. Main-reflector parameters. D is the projected aperture, F is the focal length, and DØ is the offset distance (d_0 of Figs. 2.1, 2.2, and 2.3).

LINE 8: ECENTR, FOCSUB, BETAD, ALFAD, THETED. Subreflector parameters. ECENTR is the eccentricity ($e>1$ for a hyperboloid and $0<e<1$ for an ellipsoid). FOCSUB is the interfocal distance &c. BETAD is the depression angle β. ALFAD is the angular center direction α. And THETED is the edge angle θ_E. The last four variables are defined in Figs. 2.1-2.3.

LINE 9: NELEMT. Number of feed array elements.

LINES 10, 11, ... NELEMT+9: XELEM, YELEM, ZELEM, AZIMUD, ELEVAD, REXCT, VIEXCT, PHIFOD. Array elements parameters, one line for each element. XELEM, YELEM, ZELEM, AZIMUD and ELEVAD are the element location coordinates and pointing angles, X_E, Y_E, Z_E, Az and El of Fig. 3.1, respectively. REXCT and VIEXCT are the real and imaginary parts of the excitation coefficient [EEXCIT of Eq. (3.9a)],

respectively. PHIPOD is the polarization angle ϕ_{po} of Fig 3.1 and Eq. (3.92).

Line NELEMT + 9 is the last line of the data file.

b) Output data File SORADC.OUT

All the data and output results of program SORADC are written in this data file; it is the program principal output file. It is self explanatory and the sample runs presented in this report are sufficient for its understanding.

c) Output data File PACUT1.DAT

This data file contains a table of the co-polar and cross-polar pattern values, as a function of θ_s, corresponding to the pattern cut ϕ_s = PHIS1 (see line 5 of the file SORADC.DAT). Its values are identical to the ones outputed in file SORADC.OUT. This file is intended for pattern plotting purposes, and all its gain values are in dB.

d) Output data File PACUT2.DAT

This file is similar to the previous one, with the exception that it contains the pattern values corresponding to the ϕ_s=PHIS2 cut. In the particular situation where

PHIS1 = PHIS2, this file is empty.

e) Output data file PACUT3.DAT

Again this file is similar to PACUT1.DAT except that it contains the pattern values corresponding to the ϕ_s = PHIS3 cut. When PHIS3 = PHIS2 or PHIS3 = PHIS1, this file is empty.

6.3 - Sample Runs

In the next pages we present eight complete runs of the program SORADC. They have been included in here with the sole purpose of illustrating SORADC's operation, and may not be useful engineering examples.

The first four runs demonstrate SORADC's operation when evaluating the radiation characteristics of Cassegrain and Gregorian antennas excited by a single feed element. And the last four show how to use SORADC with an array feed excited with the coefficients generated by code OFRADC [1.1]. For clarity we have given all the corresponding input and output data files, even the ones associated with code OFRAD'C.

A VAX 6340 multiprocessor computer was used for all sample runs, and the corresponding CPU time are given in order to provide an idea of SORADC's speed.

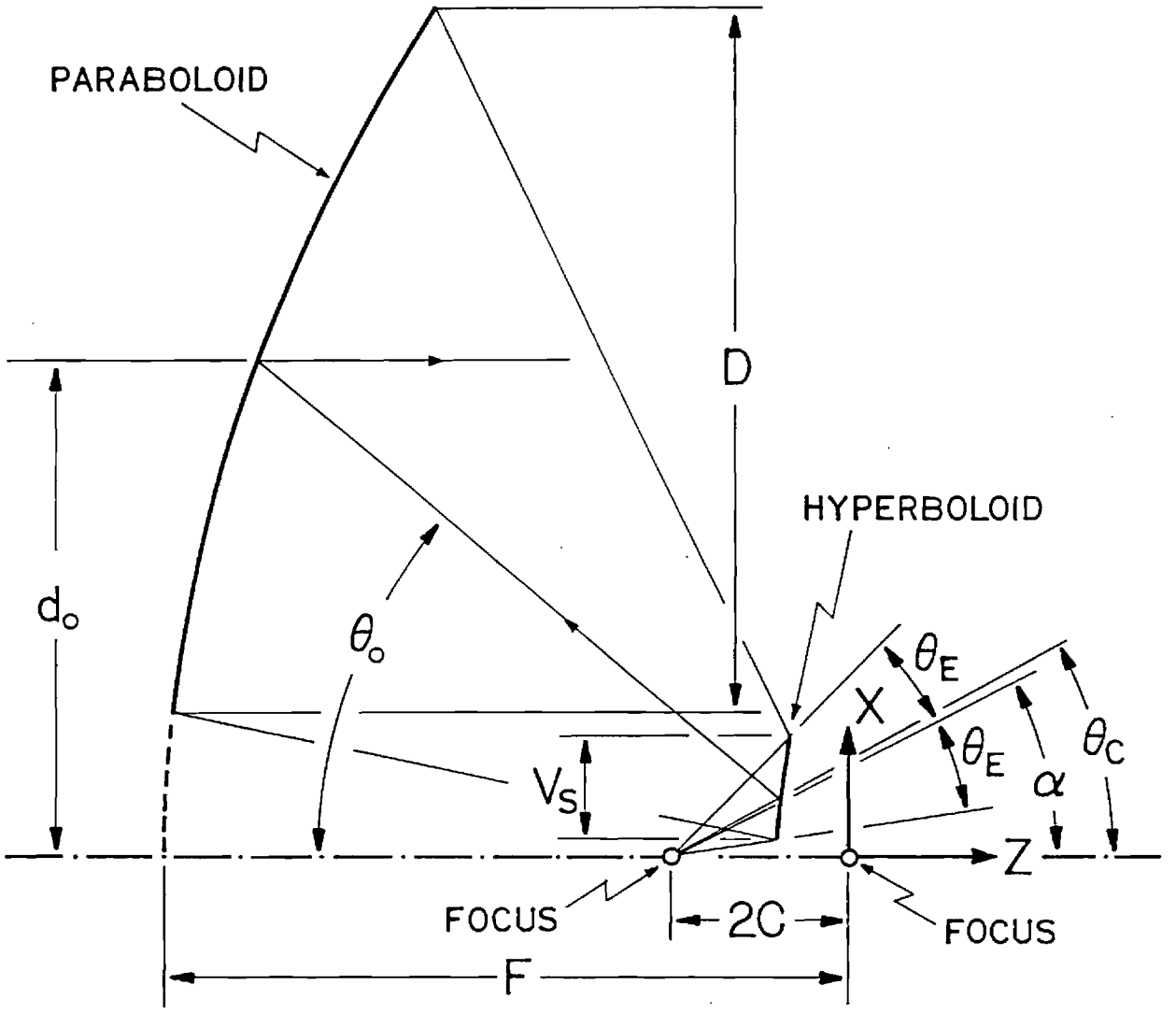
6.3.1 - Sample Run #1 : Cassegrain from the Literature. GTD Results

This first sample run corresponds to the antenna shown in Fig. 6.2. Since radiation pattern results for this system are given in [6.2], the present run provides an independent check on code SORADC operation and this is the reason for having it included in this report.

A linearly polarized raised-cosine-type feed element, pointed along the direction specified by the angle θ_c, is used to illuminate the subreflector with a pattern taper of about -10 dB in the subreflector rim direction.

The gain pattern results generated by SORADC have been ploted in Fig. 6.3 using the ploting files PACUT1, PACUT2, and PACUT3. The computations were performed using GTD on the subreflector, and the total CPU time used was about 4 minites.

In the next pages all the files used and generated by SORADC are given. The warnings present in the output file SORADC.OUT are due to the fact that for field points in the neighborhood of the main-reflector surface center, no well defined stationary phase points exist on the subreflector rim. As mentioned in Sec. 6.1, in these cases SORADC considers only subreflector reflected rays, a procedure without any noticeable accuracy degradation. We could have chosen not to print the



D = 100	F = 96
d_o = 70	θ_o = 40.062025°
β = 0°	e = 5
α = 26.592361°	θ_E = 18.647369°
2C = 25	V_S = 14.802895
	θ_C = 27.322307°

Fig. 6.2 – Cassegrain Antenna from the Literature.

WARNINGS, AND hence generate a cleaner output file, but we feel that it is always better to know exactly what a computer code is doing. IN ANY CASE, IF the user so desire, the WARNING - writing statements CAN ALWAYS be commented out INSID SUBROUTINE DFPTNF.

Comparison of Fig. 6.3 with Fig. 3 of Ref. [6.2] show that both results Agree within drawing resolution, hence validating code SORDC.

Put 15 pages
in here

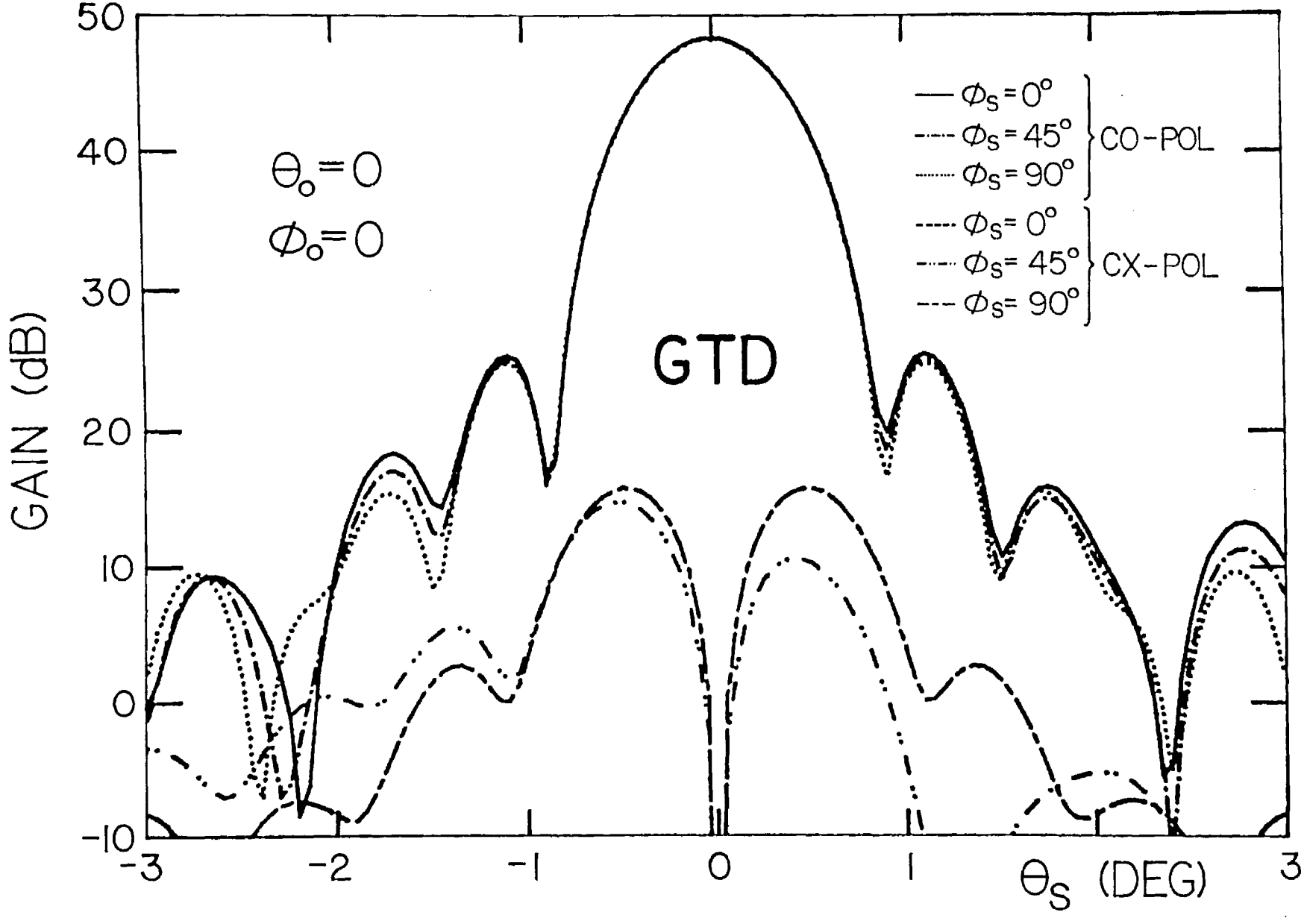


Fig. 6.3 - Cassegrain Antenna from the Literature Radiation Patterns. GTD Results

6.3.2 - Sample Run #2 : Cassegrain from the Literature. PO Results.

The antenna system of this run is identical to the one of run #1 (shown in Fig. 6.2). The only difference in here is that PO, instead of GTD, was used in the subreflector scattering computations.

The data files corresponding to this run are given in the next few pages, and Fig. 6.4 show the final radiation pattern results. As seen, Figs. 6.4 and 6.3 agree very well except at low radiation levels. This disagreement is expected and caused by the inherent differences between the PO and GTD - since GTD models better the edge diffraction mechanism its corresponding results are probably more accurate.

The main advantage of using GTD, as opposed to PO, in this particular antenna scattering computation can be easily appreciated by noting that, the PO computations took about 40 minutes o CPU time; 10 times more than required by the GTD run.

Put 10 pages IN
here

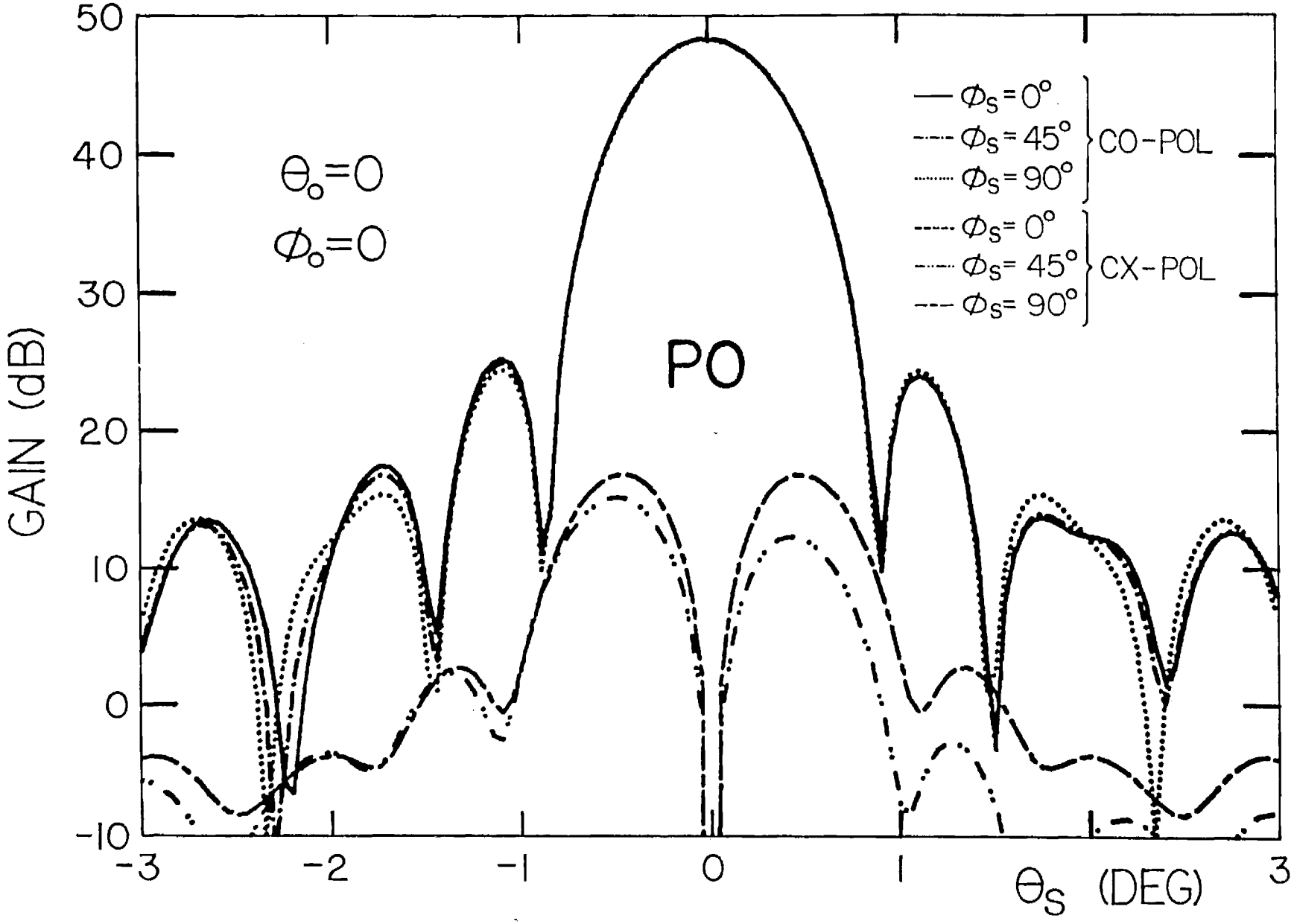
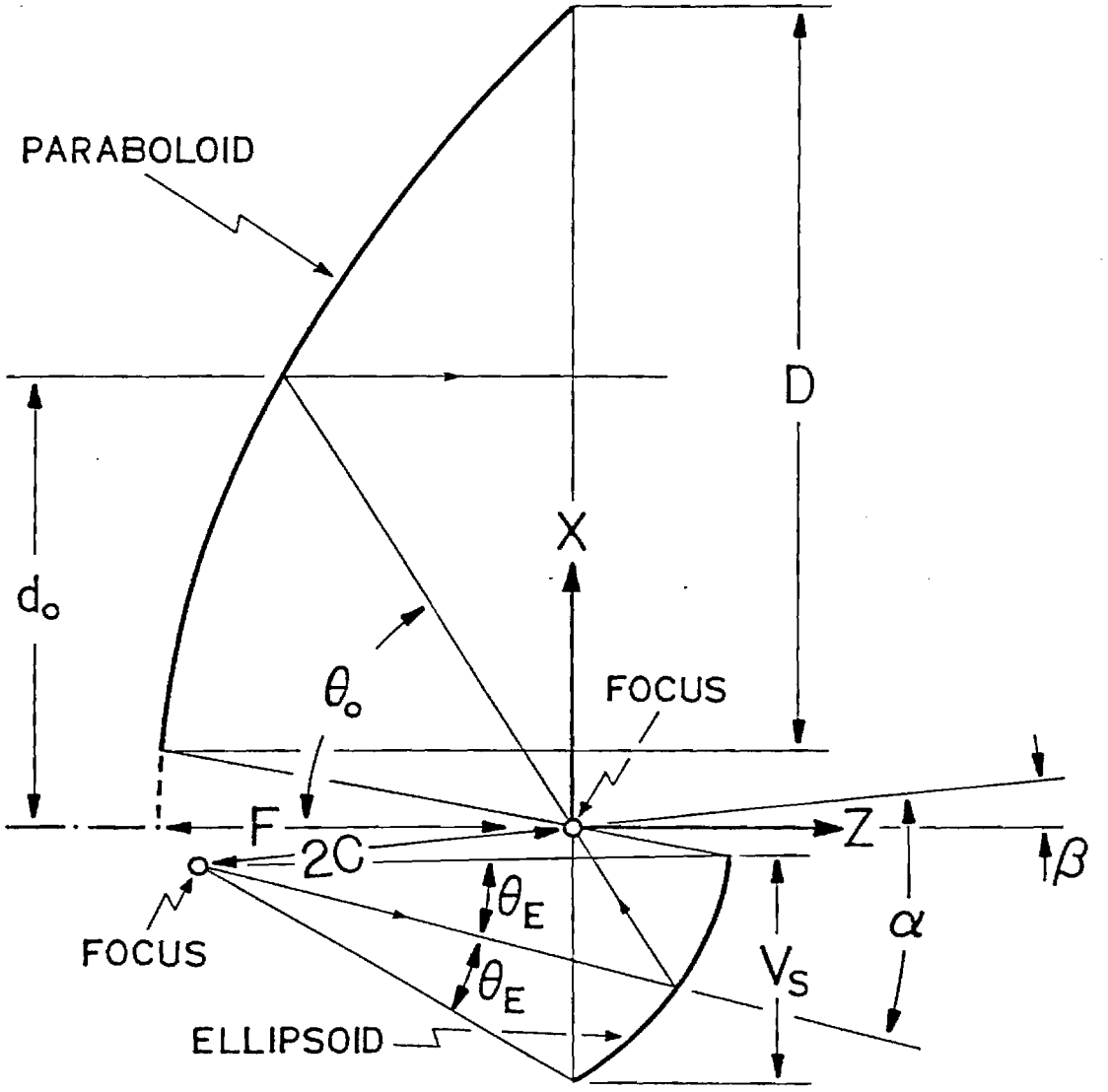


Fig. 6.4 - Cassegrain Antenna from the Literature Radiation Patterns. PO Results.

6.3.3 - Sample Run #3: Gregorian Antenna. GTD Results

The purpose of this run is to provide a test case involving a Gregorian antenna excited by a single feed element. The system geometry chosen is given in Fig. 6.5, and yield an antenna with an axially symmetric equivalent paraboloid [2.1]. A raised-cosine-type feed with -10 dB taper towards the subreflector rim was selected and aimed in the direction specified by the angle α. The data files corresponding to the SORADC run are given in the following pages, and Fig. 6.6 shows a plot of the radiation pattern.

This run took about 4.5 minutes of CPU time, half-minute more than the ammount required by sample run #1. This additional half-minute was mainly used in searching for diffraction points over the longer rim length of the Fig. 6.5 subreflector (Note that runs #3 and #1 have the same NRADAL, NRASUB, and NTARAY).



D = 100	F = 55
d_o = 60	θ_o = 57.220919°
β = 6°	e = 0.54821587
α = -20.363086°	θ_E = 15.550616°
2C = 49.889263	V_S = 30

Fig. 6.5 - Gregorian Antenna

Put 10 pages
in here

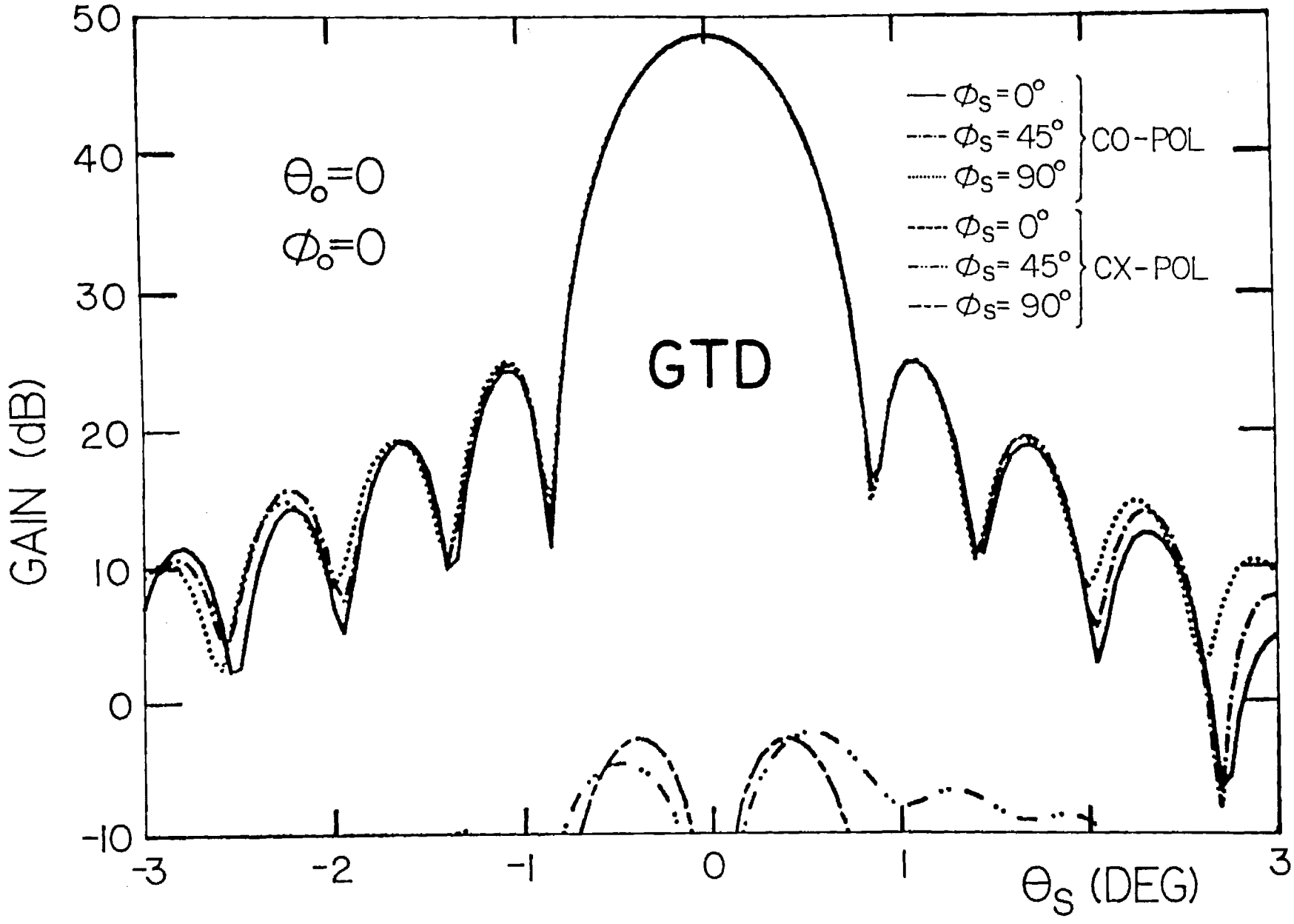


Fig. 6.6 - Gregorian Antenna Radiation Patterns. GTD Results.

6.3.4 – Sample Run #4: Gregorian Antenna. PO Results

In this run the antenna system of sample run #3, shown in Fig. 6.5, is analyzed using PO. The SORADC's data files are given ahead, and Fig. 6.7 depict the corresponding radiation patterns.

This run took about 2.5 hours of CPU time, and the agreement between Figs. 6.7 and 6.6 is much better than the one between Figs. 6.4 and 6.3. The increase in CPU time, from run #2 and the better agreement between radiation patterns, are both due the larger subreflector size used in this run.

Put 10 pages in here

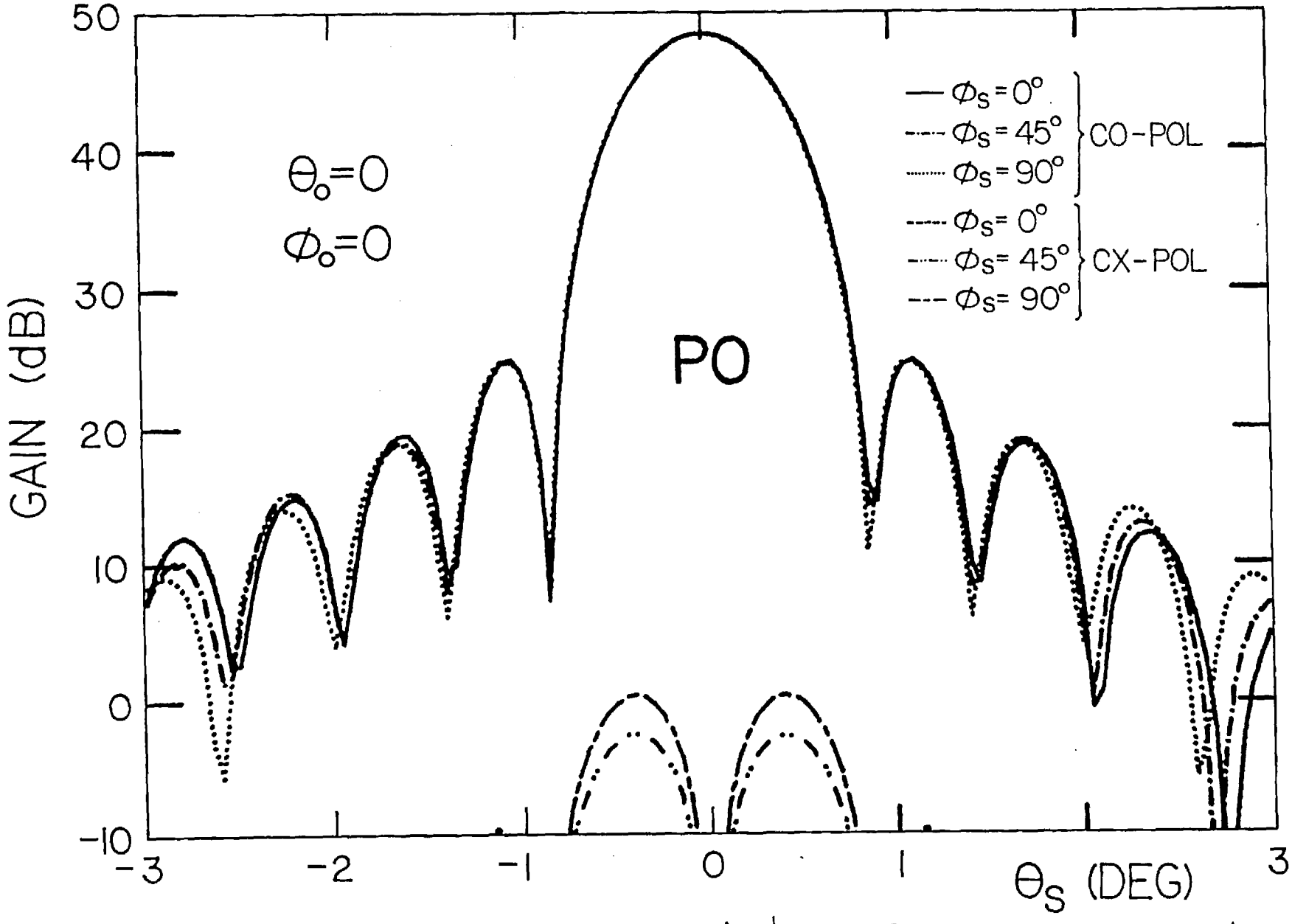


Fig. 6.7 - Gregorinn Antenna Radiation Patterns. PO Results.

6.3.5 – Sample Run #5: Cassegrain Excited by an Array Feed. GTD Results

In this run we show how to use codes SORADC and OFRADC (described in Ref. [1.1]) together. The antenna geometry is the same one used in sample run #4 of [1.1], and will not be repeated in here. Four data files are given in the following pages The first two correspond to code OFRADC, and are identical to the ones presented in [1.1] – they have been included here for completeness. And the last two are the input and output data files of code SORADC, respectively. Note that, in the file 'SORADC.DAT', all lines after the sixth have been generated automatically by code OFRADC incorporating the modification presented in Appendix I.

The SORADC's radiation pattern results are ploted in Fig. 6.8, and have been obtained using a CPU time of about three hours – this time does not include the three minutes required to run code OFRADC. We deffer comments on the radiation pattern for Sec. 6.3.6, where comparable results are obtained using PO on the sub-reflector.

Put 37 pages in here

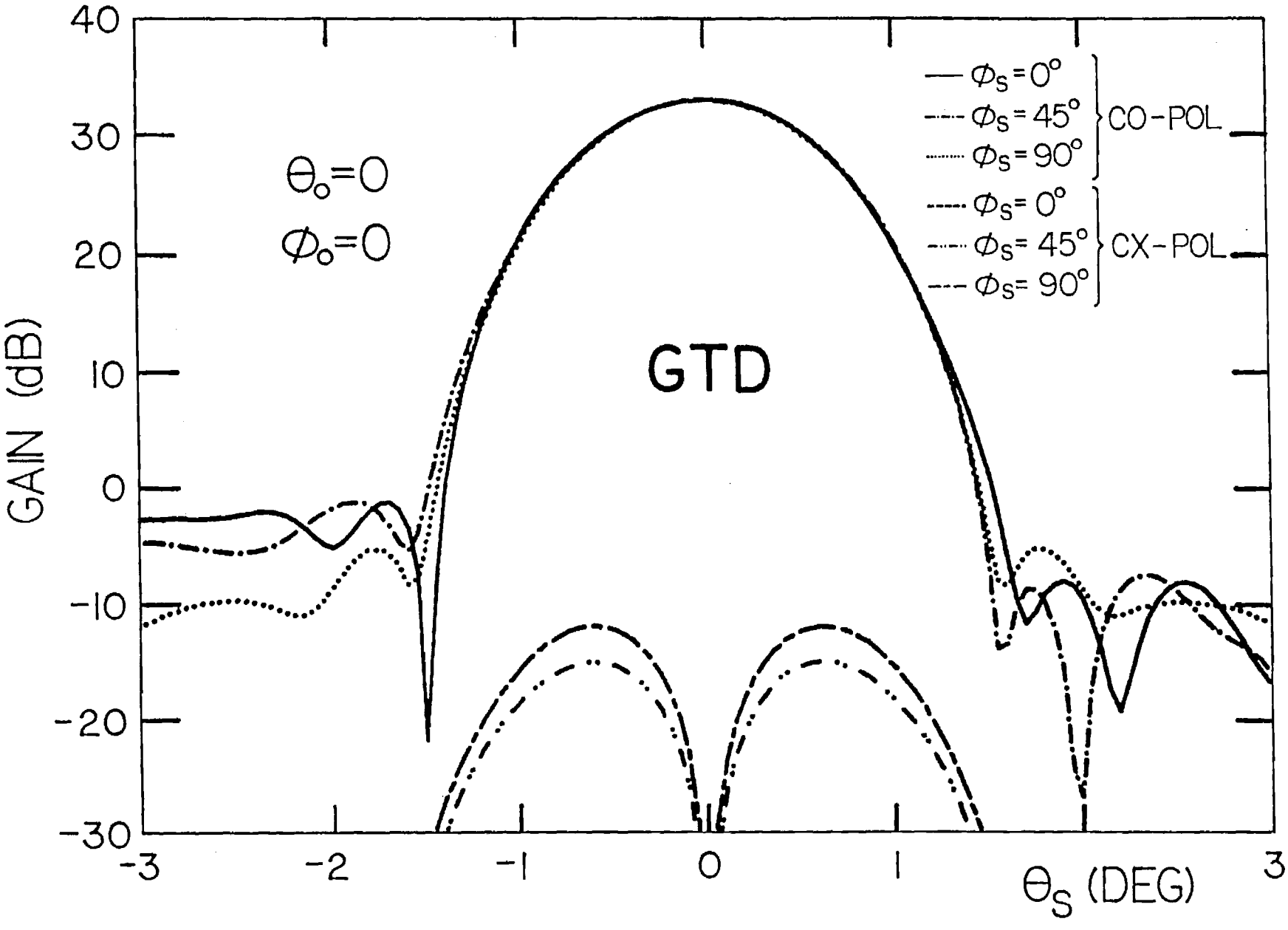


Fig. 6.8 - Radiation Pattern of Array Fed Cassegrain Antenna. GTD Results.

6.3.6 - Sample Run #6: Cassegrain Excited by an Array Feed. PO Results

This sample run corresponds to the same antenna and feed array used for sample run #5, with the difference that PO was used in the subreflector scattering evaluation. The corresponding data files are given in the following pages, and Fig. 6.9 shows a plot of the radiation pattern results.

The ammount of CPU time used by the present run was about 5 hours, two hours more than for sample run #5. Once more we experience a computational advantage of GTD over PO, this time for an array feed with 81 elements.

Comparing Figs. 6.9 and 6.8 we see a significant disagreement between the sidelobes, and also Fig 6.8 depicts a slightly assymmetric main beam. The causes of these disagreements between the PO and GTD results are unknown at this point. One possible explanation is that, since the focal region fields computed by OFRADC correspond to a tapered plane wave specified to produce zero field over the subreflector rim, slope diffraction effects [4.6] (not included in the GTD formulation used in Sec. 4.1.4) may be significant in this case. Although this remains to be confirmed, these discrepancies were not observed when more practical illuminations or larger subreflector sizes were used (see for example sample runs

#7 AND #8), AND HENCE WE FEEL THAT they do NOT CONSTITUTE A MAJOR HANDICAP OF the GTD FORMULATION.

Put 13 pages
in here

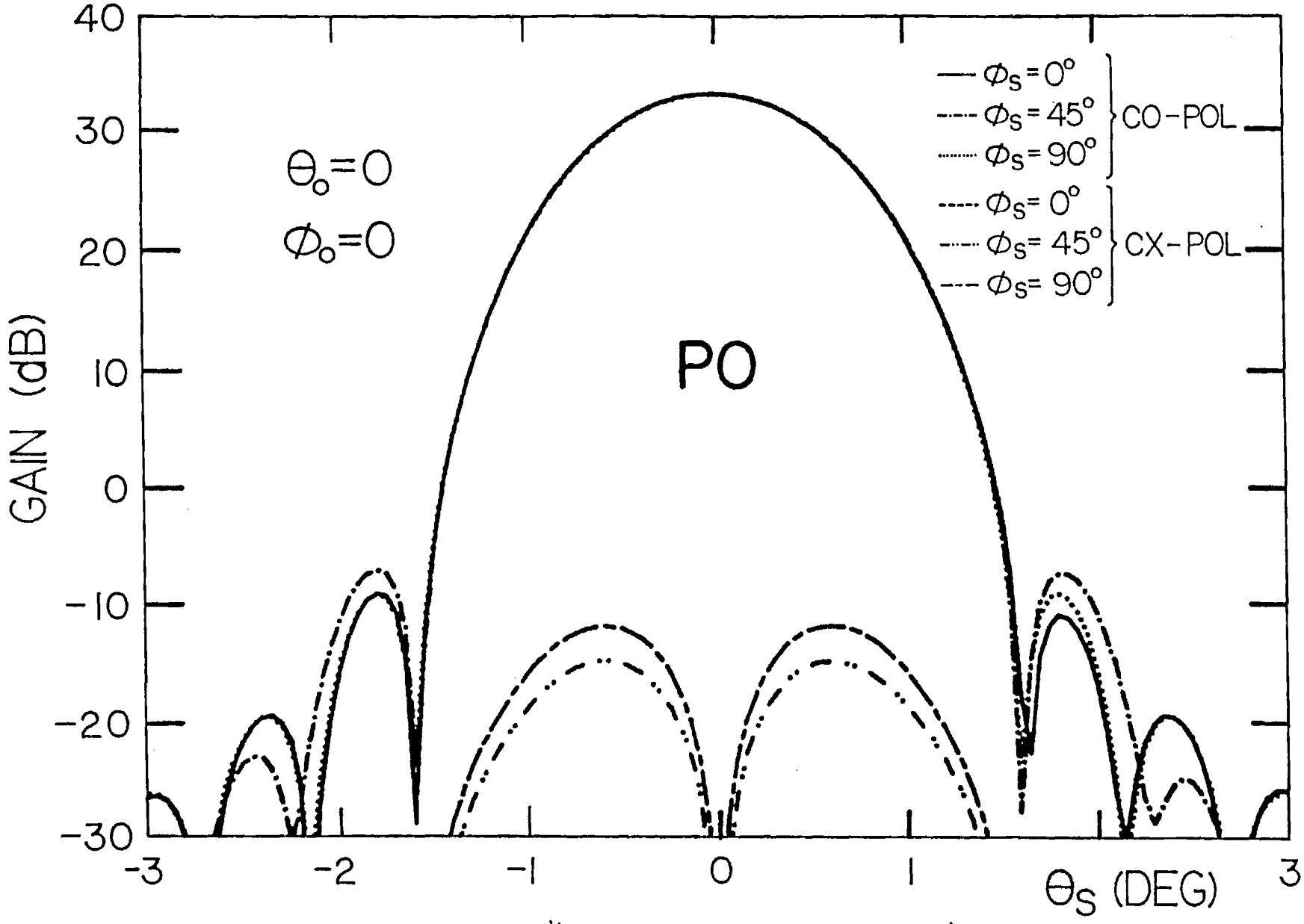


Fig. 6.9 - Radiation Pattern of Array Fed Cassegrain Antenna. PO Results.

6.3.7 - Sample Run #7 : Gregorian Excited by an Array Feed. GTD Results.

This sample run provides an example of the usage of codes OFRADC and SORADC together in conjunction with a Gregorian antenna system (code OFRADC is fully described in [1.1]). The antenna geometry is identical to the one used is sample run #6, of [1.1] — the reader is then referred to this reference for more details on the geometry. In the next pages the input and output data files of OFRADC and SORADC are given, and the final radiation pattern results are ploted in Fig. 6.10. A CPU time of about 3 minutes, and 4.5 hours, was used by the OFRADC and SORADC runs, respectively.

Put 50 pages in here

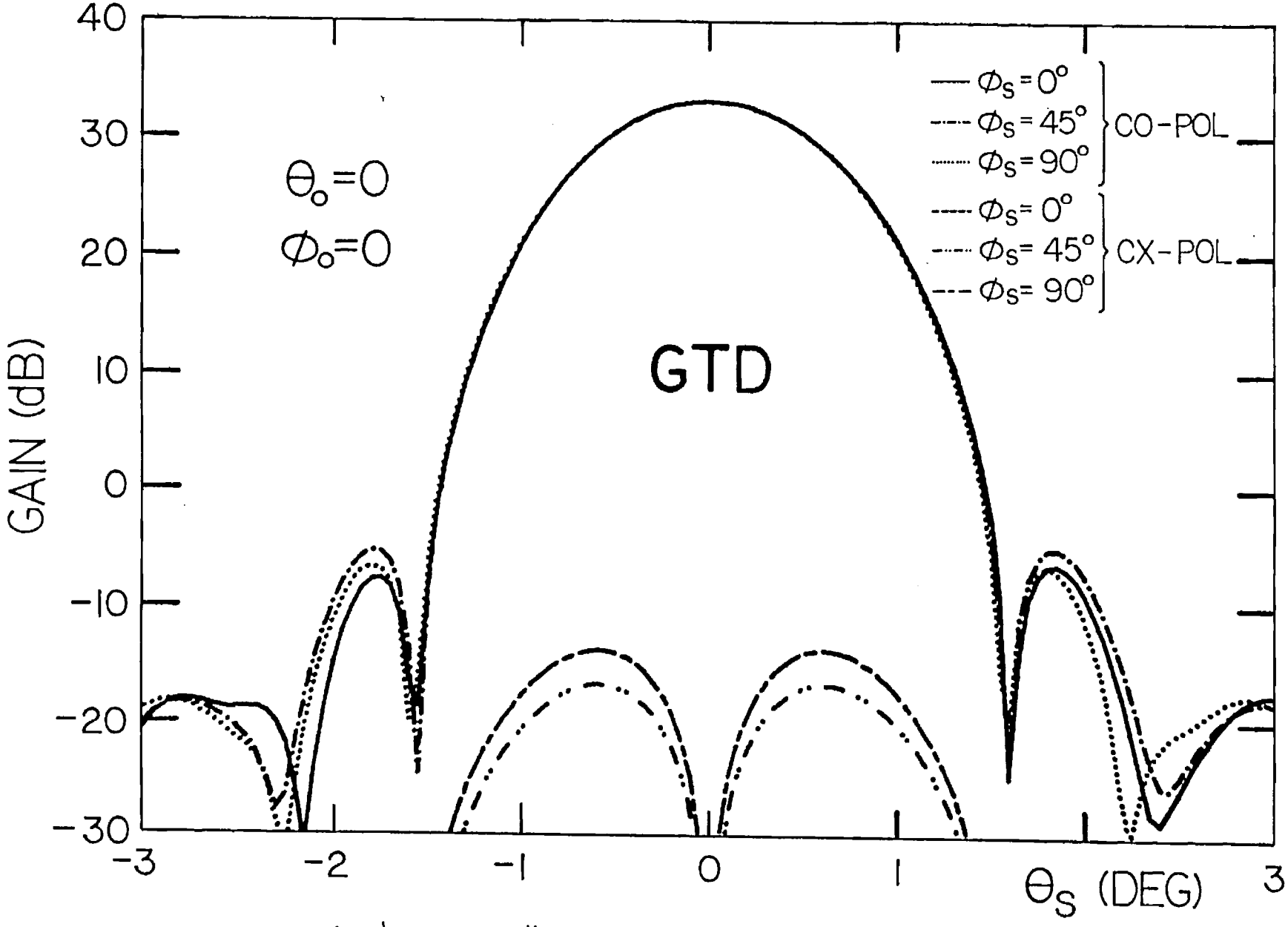


Fig. 6.10 - Radiation Pattern of Annay Fed Gregorinn Antenna. GTD Results.

6.3.2 - Sample Run #8: Gregorian Excited by an Array Feed. PO Results.

This sample run is almost identical to the previous one, the only difference is that, in here, SOLADC evaluated the sub-reflector scattering using PO. The corresponding data files are given in the next pages, and a plot of the radiation patterns is shown in Fig. 6.11. This run took about 12 hours of CPU time to complete, almost three times more than the previous one.

Comparing Figs. 6.11, 6.10, and 4.8 (of Ref. [1.1]) we observe that the agreement obtained between them is excellent. As opposed to the Cassegrain geometry used in runs # 5 and #6, the Gregorian configuration depicts a much better agreement between the PO and GTD patterns (Figs. 6.11 and 6.10). We atribute this fact to the much larger subreflector used in the Gregorian antenna.

Put 13 pages in

here

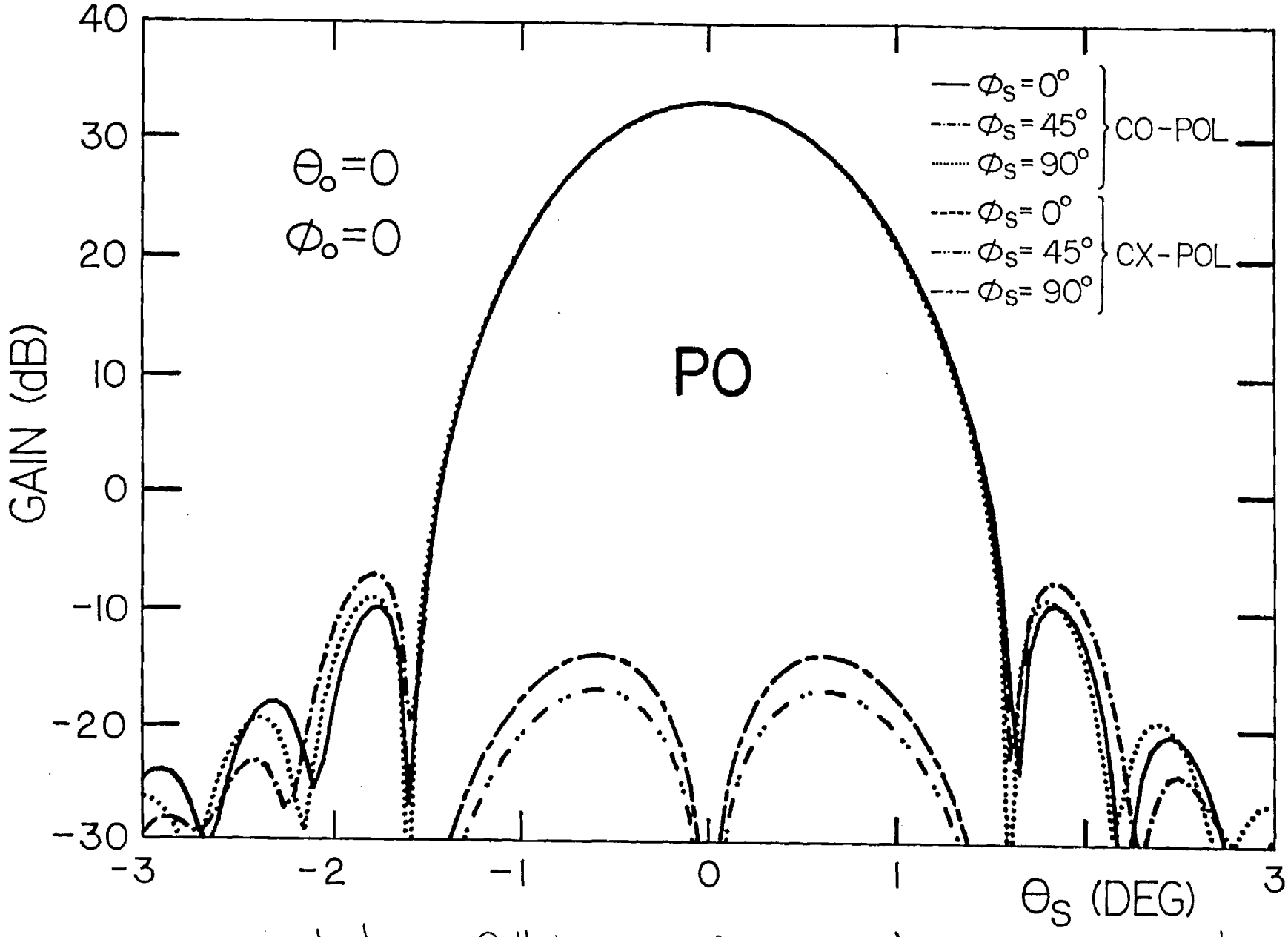


Fig. 6.11 - Radiation Pattern of Array Fed Gregorian Antenna. PO Results.

7 - Conclusion

The procedures required to evaluate the radiation characteristics of classical offset dual-reflector antennas excited by array feeds have been presented and implemented in the FORTRAN language computer program SORADC. The code SORADC is intended to be used either alone or in conjunction with code OFRADC, described in Ref. [1.1]; both codes were made fully compatible through a small upgrade introduced in OFRADC. When used with OFRADC, both codes constitute a complete package for designing and analyzing dual-reflector antennas capable of electronic beam scanning and sidelobe control.

The methods and software presented in this report have been validated through several test cases, and in particular we have included in here eight complete sample runs of code SORADC (to demonstrate its operation). All results obtained using SORADC show that this code can be successfully used to evaluate the radiation characteristics of classical offset dual-reflector antennas excited by array feeds.

8 - References

1.1 - W. V. T. Rusch and A. Prata, Jr: "Computation of Focal-region Fields of Offset Dual-Reflector Antennas for Scanning Applications," final report for RADC contract F30602-81-C-0205, task E-5-7067, #5478. Dept. of Electrical Engineering, University of Southern California, Los Angeles CA, May 1986.

1.2 - W. V. T. Rusch and P. D. Potter: "Analysis of Reflector Antennas," Academic Press, 1970.

1.3 - R. G. Kouyoumjian and P. H. Pathak: "A Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," IEEE Proc, Vol. 62, No. 11, pp. 1448-1461, Nov. 1974.

1.4 - S. W. Lee, P. Cramer, Jr, K. Woo, and Y. Rahmat-Samii: "Diffraction by an Arbitrary Sub-reflector: GTD Solution," IEEE Trans. Antennas Propagat, Vol. AP-27, No. 3, pp. 305-316, May 1979.

2.1 - W. V. T. Rusch, A. Prata, Jr, Y. Rahmat-Samii, and R. A. Shore: "Derivation and Application of the Equivalent Paraboloid for Classical Offset Cassegrain and Gregorian Antennas," in press, IEEE Trans. Antennas Propagat.

3.1 - A. C. Ludwig: "The Definition of Cross Polarization," IEEE Trans. Antennas Propagat, Vol. AP-21, No. 1, pp. 116-119, Jan. 1973.

4.1 - J. B. Keller: "Geometrical Theory of Diffraction," J. Op. Soc. Am, Vol. 52, No. 2, pp. 116-130, Feb. 1962.

4.2 - M. Born and E. Wolf: "Principles of Optics," Pergamon Press, 1983.

4.3 - W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling: "Numerical Recipes," Cambridge Univ. Press, 1986.

4.4 - D. J. Struik: "Lectures on Classical Differential Geometry," Dover Pub, Inc, 1961.

4.5 - S. W. Lee: "Differential Geometry for GTD Applications," Mathematics Notes, Note 53, Univ. of Illinois, Urbana, Oct. 1977

4.6 - J. D. Cashman and R. G. Kouyoumjian: "Comments on a Uniform Geometrical Theory of Diffraction for an Edge in a Perfectly Conducting Surface," IEEE Trans. Antennas Propagat, Vol. AP-25, No. 3, pp. 447-451, May 1977.

4.7 - J. Boersma: "Computation of Fresnel Integrals", Math. Comp, Vol. 14, No. 72, pp. 380, Oct. 1960.

4.8 - W. V. T. Rusch: "Reflector Antennas", on Numerical and Asymptotic Techniques in Electromagnetics, edited by R. Mittra, Springer-Verlag, 1975

6.1 - W. V. T. Rusch: "Antenna Notes", Vol. II, NB 846, Electromagnetics Institute,

Technical University of Denmark, Lyngby, Aug. 1974.

6.2 - Y. Rahmat-Samii: "Subreflector Extension for Improved Efficiencies in Cassegrain Antennas - GTD/PO Analysis," IEEE Trans. Antennas Propagat, Vol. AP-34, No. 10, pp. 1266 - 1269, Oct. 1986.

9 - Appendices

Appendix I - Computer Code OFRADC Upgrade

In order to make computer programs SORADC and OFRADC (described in Ref. [1.1]) fully compatible, a subroutine named OUSORA, and a corresponding call statement, was added to upgrade program OFRADC. The added code lines are shown in the next few pages.

The sole purpose of subroutine OUSORA is to partially generate code's SORADC input data file SORADC.DAT; all lines following the sixth are automatically generated by code OFRADC upgraded. These lines contain all the antenna geometry and feed array parameters (see Sec. 6.2). Of particular relevance to the GTD computations of the subreflector scattering is the fact that OUSORA puts the array feed elements, in file SORADC.DF in a sequential fashion (i.e. each feed element is near the previous and the subsequent one). This makes the process of locating subreflector reflection points very efficient, since the reflection point location corresponding to a given array element can be used as a guess when searching for the next element reflection point location.

The code lines added to upgrade OFRADC are given below.

```
C PROGRAM OFRADC.FOR
C
C THIS PROGRAM COMPUTES THE FOCAL REGION FIELDS OF AN OFFSET CASSEGRAIN
C OR GREGORIAN REFLECTOR ANTENNA, OR A PARABOLIC OFFSET REFLECTOR, WHEN
C ILLUMINATED BY A TAPERED PLANE WAVE.
C FOR THE FIELDS PRODUCED BY THE MAIN REFLECTOR, OPTION IS GIVEN TO USE
C GO+GTD (THE FAST WAY) OR PO (THE VERY SLOW WAY).
C PO SURFACE CURRENT INTEGRATION IS USED IN THE SUBREFLECTOR.
C
C INPUT DATA IS READ FROM FILE "OFRADC.DAT".
C OUTPUT IS IN FILE "OFRADC.OUT"
C OUTPUT FOR PROGRAM OFPARA.FOR IS IN FILE OFPARA.DAT
C
C A. PRATA JR. AND W.V.T. RUSCH - UNIVERSITY OF SOUTHERN
C CALIFORNIA - 1985
C
C PROGRAM UPDATED IN JULY 1989 TO GENERATE PART OF THE DATA FILE REQUIRED
C TO RUN CODE SORADC.FOR
C
C REFLECTOR PARAMETERS
      COMMON/REFPAR/D,F,DO,DS,FS,DOS
C SUBREFLECTOR PARAMETERS
      COMMON/SUBPAR/ECCNT,FOCSUB,DELTAC,BETA,THTMAX
C
      CHARACTER*80,TITLE1,TITLE2
      REAL*8 THETA,PHI,D,F,DO
      COMPLEX ETOT(3),HTOT(3)
      COMPLEX EFXG(20,20),EFYG(20,20),EFZG(20,20)
      COMPLEX HFXG(20,20),HFYG(20,20),HFZG(20,20)
      COMPLEX CEFXG
      DIMENSION EFXGA(20,20),EFXGP(20,20),EFYGA(20,20),EFYGP(20,20)
      DIMENSION EFZGA(20,20),EFZGP(20,20),HFXGA(20,20),HFXGP(20,20)
      DIMENSION HFYGA(20,20),HFYGP(20,20),HFZGA(20,20),HFZGP(20,20)
C
      OPEN(UNIT=30,FILE='OFRADC.OUT',STATUS='NEW')
C
      PI=3.1415927
      DTR=PI/180.
         .
         .
         .
         .

      WRITE(30,190) HFXG(I,J),HFYG(I,J),HFZG(I,J)
  190 FORMAT(5X,'HXG=(',E10.4,',',E10.4,')',
     *' HYG=(',E10.4,',',E10.4,')',
     *' HZG=(',E10.4,',',E10.4,')A')
C
  160 CONTINUE
C
      CLOSE(UNIT=30)
C
C WRITING THE REFLECTOR PARAMETERS AND FOCAL REGION FIELDS ON DATA FILE
C FOR PROGRAM OFPARA.FOR
C
      IF(NUMREF.EQ.1) CALL OUPARP(BETA,SHIFT,ALFA,XGSTRT,XGINC,NXG,
     *YGSTRT,YGINC,NYG,EFXG,EFYG,PHIPOL)
      IF(NUMREF.EQ.2) THEN
      CALL OUCGP(BETA,SHIFT,ALFA,XGSTRT,XGINC,NXG,YGSTRT,YGINC,NYG,
     *EFXG,EFYG,PHIPOL,ECCNT)
```

```
      CALL OUSORA(SHIFT,ALFA,XGSTRT,XGINC,NXG,YGSTRT,YGINC,NYG,
     *EFXG,EFYG,PHIPOL)
      ENDIF
C
      STOP
      END
C
C
      SUBROUTINE IMPUTD(TITLE1,TITLE2,NUMREF,NGOPO,NRADAL,NRDPAR,SHIFT,
     *ALFA,XGSTRT,XGINC,NXG,YGSTRT,YGINC,NYG,PHIPOL)
C
C THIS SUBROUTINE READS THE INPUT DATA FILE 'OFRADC.DAT', AND PUTS PARTS
C OF IT ON 'COMMON'. IT ALSO DOES BOOKKEEPING.
C
C REFLECTOR PARAMETERS
      COMMON/REFPAR/D,F,DO,DS,FS,DOS
C SUBREFLECTOR PARAMETERS
      COMMON/SUBPAR/ECCNT,FOCSUB,DELTAC,BETA,THTMAX
C INCIDENT PLANE WAVE PARAMETERS
      COMMON/INCWPA/THETA,PHI,THETAS,PHIS,HIX,HIY,ABAR,P,BBAR,Q
C
      CHARACTER*80,TITLE1,TITLE2
      REAL*8 THETA,PHI,D,F,DO
         .
         .
         .
         .
C
C WRITING THE CONJUGATE OF THE FOCAL FIELDS, THE FIELD EXCITATION
C COEFFICIENTS FOR PROGRAM "OFPARA.FOR"
C
      DO 10 I=1,NXG
      DO 10 J=1,NYG
      CEFXG=CONJG(EFXG(I,J)*COSPOL+EFYG(I,J)*SINPOL)
      CREFXG= REAL(CEFXG)
      CIEFXG=AIMAG(CEFXG)
      WRITE(40,*) CREFXG,CIEFXG
   10 CONTINUE
C
      CLOSE(UNIT=40)
C
      RETURN
      END
C
C
      SUBROUTINE OUSORA(SHIFT,ALFA,XGSTRT,XGINC,NXG,YGSTRT,YGINC,
     *NYG,EFXG,EFYG,PHIPOL)
C
C THIS SUBROUTINE PARTIALLY CREATES THE FILE "SORADC.DAT", THE INPUT DATA
C FILE REQUIRED TO RUN CODE SORADC.
C
C
C REFLECTOR PARAMETERS
      COMMON/REFPAR/D,F,DO,DS,FS,DOS
C SUBREFLECTOR PARAMETERS
      COMMON/SUBPAR/ECCNT,FOCSUB,DELTAC,BETA,THTMAX
C
      REAL*8 D,F,DO
      COMPLEX EFXG(20,20),EFYG(20,20),CEFXG
C
```

```
      OPEN(UNIT=50,FILE='SORADC.DAT',STATUS='NEW')
C
C SOME USEFUL RESULTS
C
      PI=3.1415927
      RTD=180./PI
      ECENTR=ABS(ECCNT)
      OOMAGN=(ECENTR-1.)/(ECENTR+1.)
      TWOCMS=FOCSUB-SHIFT
      PHIPOD=PHIPOL*RTD
      SINPOL=SIN(PHIPOL)
      COSPOL=COS(PHIPOL)
C
      SINBE=SIN(BETA)
      COSBE=COS(BETA)
      SINBMA=SIN(BETA-ALFA)
      COSBMA=COS(BETA-ALFA)
C
      AZIMUT=0.
      ELEVAT=-(ALFA-BETA)*RTD
C
C COMPUTING THE SUBREFLECTOR CENTER ANGLE AND ANGULAR EXTENSION, AS SEEN
C FROM THE SYSTEM FOCUS
C
      IF(ECCNT.GE.0.) THEN
      BOTANG=2.*ATAN(OOMAGN*TAN((PI-DELTAC-THTMAX)/2.))
      TOPANG=2.*ATAN(OOMAGN*TAN((PI-DELTAC+THTMAX)/2.))
      ALFAC =(TOPANG+BOTANG)/2.
      THETAE=(TOPANG-BOTANG)/2.
      ELSE
      BOTANG=-2.*ATAN(OOMAGN*TAN((DELTAC-THTMAX)/2.))
      TOPANG=-2.*ATAN(OOMAGN*TAN((DELTAC+THTMAX)/2.))
      ALFAC =(TOPANG+BOTANG)/2.
      THETAE=(TOPANG-BOTANG)/2.
      ENDIF
C
C WRITING THE REFLECTOR GEOMETRIC PARAMETERS
C
      WRITE(50,*) DS,FS,DOS
C
C WRITING THE SUBREFLECTOR GEOMETRIC PARAMETERS
C
      ALFACD=ALFAC*RTD
      BETAD=BETA*RTD
      THETED=THETAE*RTD
      WRITE(50,*) ECENTR,FOCSUB,BETAD,ALFACD,THETED
C
C WRITING THE NUMBER OF FOCAL GRIN POINTS
C
      NELEMT=NXG*NYG
      WRITE(50,*) NELEMT
C
C WRITING THE FIELD POINT POSITIONS AND CORRESPONDING CONJUGATE OF THE FOCAL
C FIELDS--THE ARRAY ELEMENTS POSITION AND FIELD EXCITATION COEFFICIENTS FOR
C PROGRAM "SORADC.FOR". SPECIAL CARE IS TAKEN TO ASSURE THAT EACH ELEMENT
C POSITION IS NEXT TO THE PREVIOUS ONE, TO ASSURE THAT PROGRAM SORADC WILL
C HAVE THE BEST POSSIBLE EFFICIENCY WHEN USING GTD OR GO
C
      DO 10 I=1,NXG
      XGRID=XGSTRT+XGINC*FLOAT(I-1)
```

```
C
      DO 10 J=1,NYG
      IF((I/2)*2.NE.I) THEN
      YGRID=YGSTRT+YGINC*FLOAT(J-1)
      ELSE
      YGRID=YGSTRT+YGINC*FLOAT(NYG-J)
      ENDIF
C
C GETING THE FEED POSITION IN THE PARABOLIC REFLECTOR CARTESIAN COORDINATES
C
      XFIELD= XGRID*COSBMA-TWOCMS*SINBE
      YFIELD= YGRID
      ZFIELD=-XGRID*SINBMA-TWOCMS*COSBE
C
C GETING THE EXCITATION COEFFICIENTS AND WRITING ALL PARAMETERS IN THE DATA
C FILE
C
      CEFXG=CONJG(EFXG(I,J)*COSPOL+EFYG(I,J)*SINPOL)
      CREFXG= REAL(CEFXG)
      CIEFXG=AIMAG(CEFXG)
      WRITE(50,*) XFIELD,YFIELD,ZFIELD,AZIMUT,ELEVAT,CREFXG,CIEFXG,
     *PHIPOD
   10 CONTINUE
C
      CLOSE(UNIT=50)
C
      RETURN
      END
C
C
      SUBROUTINE OUCGP(BETA,SHIFT,ALFA,XGSTRT,XGINC,NXG,YGSTRT,YGINC,
     *NYG,EFXG,EFYG,PHIPOL,ECCNT)
C
C THIS SUBROUTINE CREATES THE FILE "OFPARA.DAT", WHICH HAS THE CASSEGRAIN
C /GREGORIAN EQUIVALENT PARABOLIC REFLECTOR PARAMETERS AND FOCAL REGION
C FEEDS.
C
C "OFPARA.DAT" IS THE IMPUT DATA FILE FOR THE PROGRAM "OFPARA.DAT", WHICH
C COMPUTES THE FAR FIELDS RADIATED BY A PLANAR ARRAY OF FEEDS WITH
C ARBITRARY POSITION AND ORIENTATION.
C NOTE THAT THIS SUBROUTINE DOESN'T WRITE THE FILE OFPARA.DAT COMPLETELLY.
C SOME LINES STILL MUST BE ADDED LATER BY THE USER OF OFPARA.
C
C REFLECTOR PARAMETERS
      COMMON/REFPAR/D,F,DO,DS,FS,DOS
C EQUIVALENT PARABOLOID PARAMETERS
      COMMON/EQPARP/FEQ,ALFAEQ,DOEQ
C
      REAL*8 D,F,DO
      COMPLEX EFXG(20,20),EFYG(20,20),CEFXG
         .
         .
         .
         .
```

Appendix II - Computer Code SORADC.FOR

Magnetospectral analysis of tunneling processes in a double-quantum-well tunneling structure

J. J. L. Rascol, K. P. Martin, S. Ben Amor,* and R. J. Higgins
School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

A. Celeste and J. C. Portal
Institut National des Sciences Appliquées, 31077 Toulouse, France and Service National des Champs Intenses, Centre National de la Recherche Scientifique, 38042 Grenoble, France

A. Torabi, H. M. Harris, and C. J. Summers
Georgia Tech Research Institute and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 6 February 1989; revised manuscript received 24 August 1989)

Using high magnetic fields to tune the Landau-level separation through phonon energies, we identify the off-selection-rule tunneling processes in an $Al_xGa_{1-x}As/GaAs$ double-quantum-well tunneling structure. Different elastic and inelastic tunneling mechanisms are identified. The effective mass of the tunneling electrons is directly determined from the crossing point of the elastic and inelastic processes. We observe the field-induced suppression of the elastically scattered contribution to the peak tunneling current. An upper limit on the resonant part of the tunneling current is established.

MS code no. BBR418B 1990 PACS number(s): 72.20.My, 73.40.Gk, 73.20.Dx

It is ordinarily assumed that in quantum-well tunneling structures (QWT's), tunneling conserves energy and momentum transverse to the tunneling direction ($k_\perp$). However, the experimental observation of a nonzero valley current in QWTS's directly shows that tunneling via selection-rule-disobeying processes takes place as well.

Studies show that elastically scattered electrons can contribute to the tunneling current.[1,2] In addition, features in the valley-current region of QWTS's (where *only* selection-rule-violating tunneling takes place) have been attributed to inelastic tunneling from the emitter into the well involving longitudinal-optical- (LO) phonon emission, or to (quasi-) elastic tunneling processes.[3–6] Tunneling processes can thus be classified as (quasi-) elastic and inelastic mechanisms with or without $k_\perp$ conservation.

This work shows, though our experimental results and a simple description, which different selection-rule-obeying and selection-rule-disobeying tunneling mechanisms contribute to different portions of the current-voltage (I-V) curves. A double quantum-well tunneling structure (DQWTS) was used to create a tunneling current (I_T) going from a two-dimensional electron gas (2D EG) in one well to a 2D EG in the other, giving well-defined initial and final states compared to a broad 3D emitter energy range.[7] Our results show that at the external bias (V_b) equal to the peak bias voltage (V_p), the magnetic field suppresses the $k_\perp$-nonconserving portion of I_T. The relative amount of non-selection-rule-obeying tunneling can be determined at V_p where ideally only selection-rule obeying tunneling should occur. This permits us to put an upper limit on the resonant (coherent) tunneling sought after in QWTS's.[8–10]

In the valley-current regime there is a B-induced enhancement of elastic-scattering-mediated tunneling at certain biases and magnetic field. We identify valley-current features arising from elastic $k_\perp$-nonconserving tunneling, and inelastic tunneling (for both $k_\perp$-conserving and $k_\perp$-nonconserving phonon-emission processes). The analysis of selection-rule disobeying tunneling is obtained from the B dependence of a feature's bias position. An effective tunneling mass is determined directly from the data *without* the use of a model describing voltage drops across the structure.[2,11]

These experiments were conducted in the range $0 < B < 30.5$ T, and are to our knowledge the most detailed magnetotunneling studies up to these high fields. At these high fields the cyclotron energy (up to 53 meV in GaAs at $B = 30.5$ T) is much larger than the LO-phonon energies, permitting us to make a careful magnetospectral analysis of the different tunneling mechanisms.

The DQWTS had well widths chosen so the lowest bound levels align at only one bias voltage (Fig. 1) at which tunneling occurs (under conservation of energy and $k_\perp$). The structure was grown by molecular-beam epitaxy and consisted of the following layers: an n^+-type GaAs substrate, a doped 25-period (14 Å/period) $Al_xGa_{1-x}As/GaAs$ superlattice, a 0.7-μM GaAs layer with graded doping from 5×10^{17} to 1×10^{17} cm^{-3}, a 51-Å layer of undoped GaAs, the DQWTS, a 51-Å undoped GaAs layer, and 0.9 μm of GaAs doped from 1×10^{17} to 5×10^{-3}. The DQWTS consists of 51 Å of undoped $Al_{0.35}Ga_{0.65}As$, 59 Å undoped GaAs, 51 Å of undoped $Al_{0.35}Ga_{0.65}As$, 28 Å of undoped GaAs, and 51 Å of undoped $Al_xGa_{1-x}As$. Additional details on the growth and processing are reported in Ref. 7. The material was

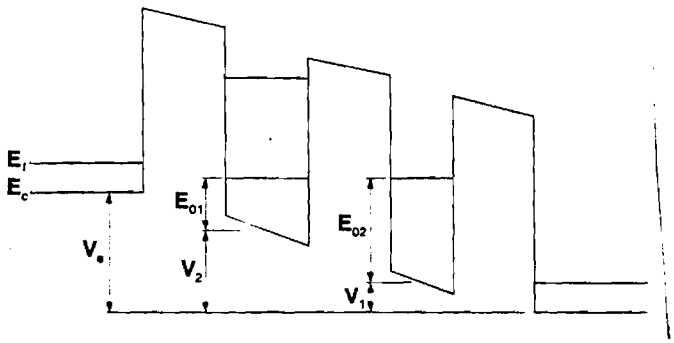


FIG. 1. A schematic representation of the GaAs/$Ga_{1-x}Al_xAs$ DQWTS conduction-band profile at peak bias voltage.

fabricated into 25-μm-diam devices and measurements were conducted at 4.2 K, with B parallel to the tunneling direction.

Figure 2 shows the strong magnetic field dependence of the I-V characteristics of the sample. For fixed $V_b < V_p$, I_T decreases with increasing magnetic field. At $B=0$, V_p occurs at 260 mV and remains essentially unchanged for B up to 30.5 T. On the other hand (Fig. 3), the main peak amplitude (I_p) decreases with B, resulting in an overall change of $\sim 40\%$.

A self-consistent calculation including the emitter and collector electrode doping profiles shows a maximum in the tunneling current at $250 < V_b < 270$ mV.[12] The calculated conduction profile and Fermi energy indicate the coexistence of 2D and 3D electrons in the emitter. This is in contrast to structures with either lightly doped emitters or wide spacer layers where the potential bump in the emitter results in a quasi-2D electron population in the accumulation layer.[6]

The valley-current regime $V_b > V_p$ (Fig. 2) exhibits distinct features whose position and amplitude are B dependent, arising from a combination of elastic and inelastic scattering processes. At $B=0$ the most prominent I_v peak (labeled LO_0) appears at $V_b=360$ mV. For $B>0$, this becomes better defined and is accompanied by *decreases* in the valley current on either side. At $B>10$ T, a second peak, LO_1 appears to the right of LO_0. Additional peaks (LO_i, $i=2$, 4, and 6) develop at higher V_b [inset of Fig. 2(a)]. Figure 4 shows their position as a function of magnetic field. V_p is fixed for B up to 30.5 T, and the position of LO_0 is independent of magnetic field for $B<21$ T. For $B>21$ T the LO_0 bias voltage increases linearly with B and these upper-field data points extrapolate to V_p. On the other hand, the positions of LO_i increase linearly with B from zero field. By extrapolating their positions in the limit of $B \to 0$, one sees they intercept the LO_0 bias voltage. For some of the valley-current peaks, the magnetic field enhancement is sufficient to destabilize their negative-differential-resistance regimes [Fig. 2(a)].

Tunneling depends on the matchup of an electron's allowed energy and momentum between both parts of a DQWTS (e.g., from the emitter to first well or first well to the second). At $B=0$ the energy of a 3D electron in the emitter electrode is

$$E_e = V_e + \hbar^2/(2m^*)(k_{e2}^2 + k_{e1}^2)\ , \tag{1}$$

with respect to the collector. The energy E_i, in the first ($i=1$) or second well ($i=2$) is

$$E_i = V_i + E_{0i} + \hbar^2/(2m^*)k_{i1}^2\ . \tag{2}$$

V_e and V_i are, respectively, the voltage differences between the collector and the emitter or ith well from the externally applied bias. The momentum has longitudinal

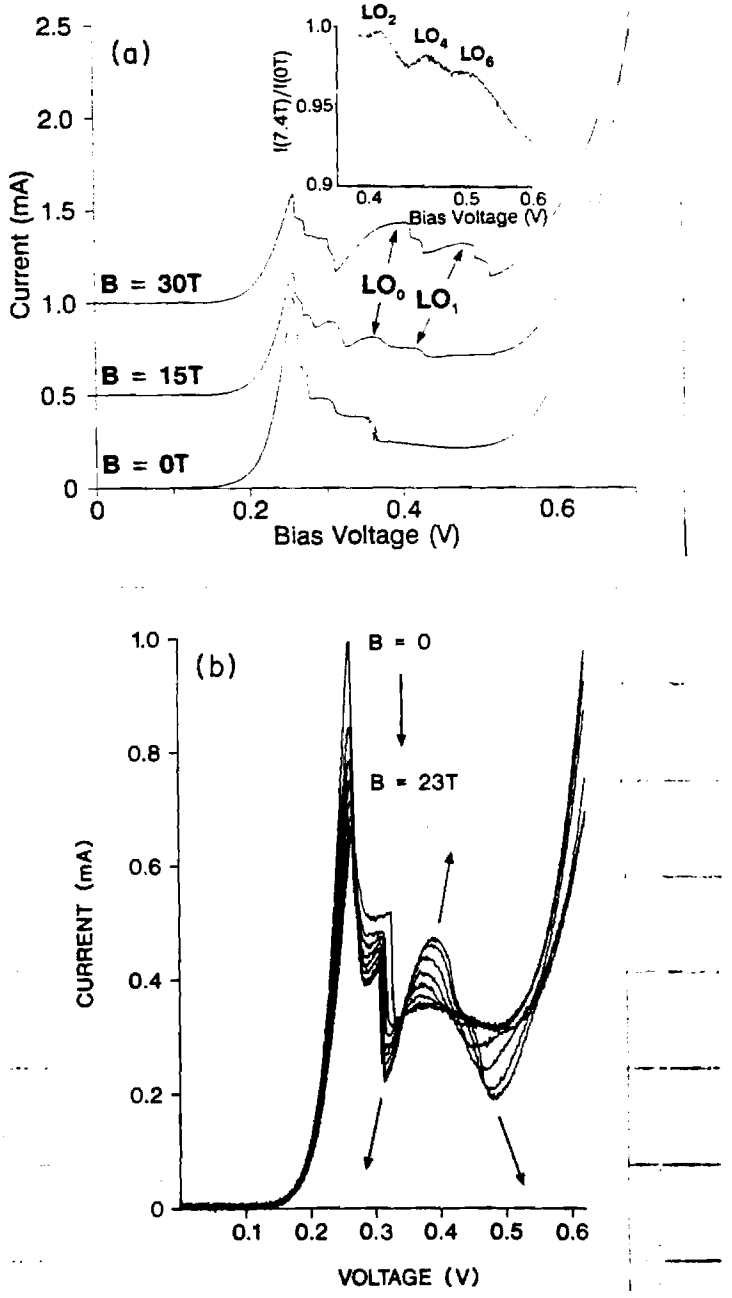


FIG. 2. (a) The current-voltage characteristics of the DQWTS at 4.2 K in magnetic fields ($\mathbf{B} \| \mathbf{J}$) up to 30.5 T. The inset shows the ratio between the tunneling current at $B=7.4$ T and zero field. (b) I-V curves for $0 < B < 23$ T for another device on the same wafer. The arrows denote the direction of the B-induced changes of different portions of the I-V curve.

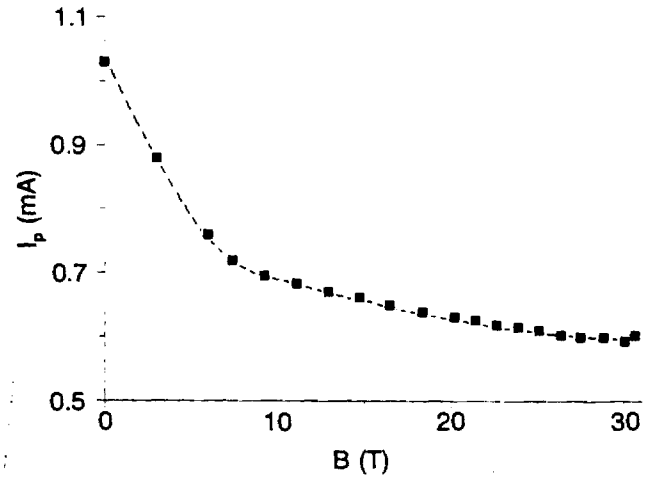


FIG. 3. The magnetic-field-induced decrease of the peak tunneling current.

(k_{ez}) and transverse ($k_{e\perp}$) components in the emitter and only transverse components ($k_{i\perp}$) in the wells. E_{0i} is the energy of the bound level in the ith well.

For $B>0$, the density of states is partitioned into Landau levels (LL's) in the emitter (3D like) and the two wells (2D like), and the transverse momentum is quantized. The energy spectrum in Eqs. (1) and (2) is changed by replacing transverse-momentum terms with $\hbar\omega_c(n_i+\frac{1}{2})$. The cyclotron frequency is $\omega_c=eB/(m^*)$, where m^* is the effective mass, and n_i is the Landau-level index for the emitter or ith well.

At $V_b=V_p$ (Fig. 1), the emitter conduction band and bound levels in the wells are aligned ($V_c=V_1+E_{01}=V_2+E_{02}$). Only electrons with $k_{ez}=0$ obey the tunneling selection rules of conservation of energy and $k_\perp$ (or Landau index when $B>0$). Rigorously speaking, electrons with $k_{ez}=0$ are unable to tunnel. However, a relaxation of this constraint occurs if they are elastically scattered into a higher-transverse-momentum state such that $k_z\to 0$. Elastic scattering by ionized impurities,[14] or surface roughness,[13] has been suggested to cause this transition. If this is permitted, then I_p is composed of two parts: one from $k_{ez}=0$ electrons which tunneling from a state of given energy and $k_{e\perp}$ into an *identical* energy and transverse-momentum state, and the second, $k_{e\perp}$-nonconserving part of I_p that comes from $k_{ez}=0$ electrons that must be elastically scattered into an allowed energy and $k_\perp$ state.

When $B=0$, the dispersion relations are continuous, allowing (under the relaxed conditions of nonconservation of $k_\perp$) any $k_{ez}=0$ electron in the emitter to tunnel. However, for $B>0$, the gaps that appear in the energy spectrum *exclude* some elastically scattered emitter electrons from tunneling. The excluded electrons are those *not* satisfying the condition

$$\hbar^2/(2m^*)k_{ez}^2=\hbar\omega_c(n_1-n_e),\quad n_1>n_e\ . \tag{3}$$

Even though Landau levels have nonzero widths, as B increases, the gaps in the density of states widen, eliminating more $k_{ez}=0$ electrons from I_T.

The B dependence of I_p (Fig. 3) shows a strong decrease that approaches saturation above ~26 T, with an overall change $\Delta I_p=-40\%$. The saturation of ΔI_p is consistent within a picture of B-induced suppression of $k_\perp$-nonconserving tunneling. When all elastically scattered electrons from $k_{ez}=0$ states have been excluded, only $k_{ez}=0$ electrons [which are unaffected by B (Ref. 14)] tunnel and I_p is constant for further increases of B. From this, we determine the amount of current coming from $k_\perp$-*non*conserving electrons. In this DQWTS, we see that at V_p more than one-third of I_p tunnels this way.[15]

Simple I-V measurements cannot discriminate between sequential and truly resonant (coherent) parts of the tunneling current.[9,10] However, the above analysis permits an upper limit to be put on the amount of resonant tunneling (in our case, less than two-thirds). This establishes the size of the elastically scattered portion of I_p at $B=0$, which cannot tunnel resonantly but must get through sequentially.

We have observed this large decrease in I_p in other QWTS's with narrow spacer layers. In contrast, we[15] and others[6,16] see only a small effect of B on I_p for QWTS's with wide spacer layers or lightly doped emitters. This different behavior is consistent with the idea that for narrow spacers there are 2D and 3D emitters with $k_{ez}\neq 0$ electrons, and for wide spacer layers or lightly doped emitters there is a quasi 2D accumulation (and $k_{ez}=0$ electrons) in the emitter.

The reduction of scattering by lowering the emitter doping or widening the spacer layer[17] provides an alternative (but not necessarily separate) mechanism for this different behavior in I_p. An increase in scattering in the emitter (for narrower spacer layers) would increase the emitter LL width. As a result, more emitter electrons will be in energy-momentum states that correspond to gaps between the LL's that form in the quantum well. The emitter electrons that line up with those gaps are

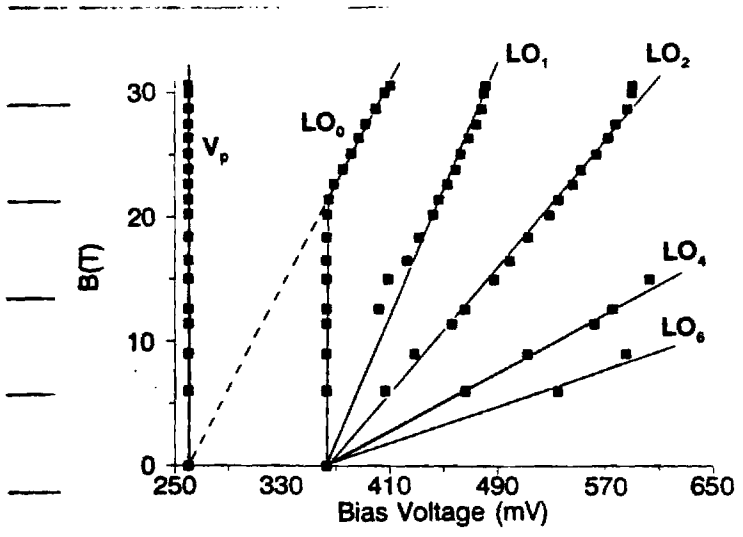


FIG. 4. Magnetic field vs bias fan chart of the peak current and LO-phonon-related peaks.

thus inhibited from tunneling, resulting in a reduction in I_p. The expected behavior from this mechanism is similar to the one previously proposed. The importance of each of these mechanisms could be understood by a systematic study of QWTS's with different emitters but similar barriers and wells.

For $V_b > V_p$, the energies of electrons in the emitter and wells can still be described by Eqs. (1) and (2) [or by the replacement of $\hbar^2/(2m^*)k_{i\perp}^2$ with $\hbar\omega_c(n_i+\frac{1}{2})$ if $B>0$]. However, the electrons do not obey the tunneling selection rules described for $V_b < V_p$. Rather, the energy and momentum relation between the emitter and well 1 for elastic (or inelastic due to emission of an LO phonon with energy $\hbar\omega_{LO}$) tunneling is

$$\Delta E_{e\text{-}1}+\hbar^2/(2m^*)k_{cz}^2 = \begin{cases} \hbar^2/(2m^*)(k_{1\perp}^2-k_{e\perp}^2)(+\hbar\omega_{LO}), & B=0 \quad (4) \\ \hbar\omega_c\Delta n\,(+\hbar\omega_{LO}), & B>0 \quad (5) \end{cases}$$

where $\Delta E_{e\text{-}1}=V_e-V_1+E_{01}$ and $\Delta n=n_1-n_e$. Similar expressions, but with $\Delta E_{e\text{-}1}\rightarrow\Delta E_{1\text{-}2}=V_1-V_2+E_{01}-E_{02}$, $\Delta n=n_1-n_2$, and without a longitudinal-momentum term, describe tunneling between the first and second wells. The total Landau-index variation across the whole structure is $\Delta n_T=n_2-n_e$.

For elastic processes the energy is conserved and the electron tunnels into a higher $k_\perp$ state. For $B>0$, this occurs if V_b and B align the Landau levels, thus satisfying Eq. (5). For each Δn_T, there will be a peak in current whose bias-voltage position moves linearly with B, and whose slope $\Delta V_b/\Delta B \propto \Delta n_T$. In the limit of $B\rightarrow 0$ their voltage position extrapolates to V_p. When Eq. (5) is not satisfied, there is no alignment between Landau ladders and the elastic scattering component of the valley current is suppressed.

Inelastic tunneling via LO-phonon emission has been demonstrated by the appearance of replica peaks in the valley-current regimes of both single and double QWTS's.[4,5] When $k_\perp$ is conserved, this process is maximized when V_b sets the left-hand side of Eqs. (4) and (5) equal to $\hbar\omega_{LO}$. Since $\Delta n_T=0$, the bias voltage position of this phonon-replica peak will be independent of B.

Peak LO_0 (Fig. 4) at 360 mV does not move for $B<21$ T, which shows that it is due to inelastic $k_\perp$-conserving tunneling. The bias difference between LO_0 and the main peak is 100 mV ($\gg\hbar\omega_{LO}$), while V_p is in good agreement with the predicted position. A simple description of the voltage drops across the sample is obtained using Eq. (5) and the bias-voltage slopes of the LO_i features. This shows that a GaAs LO phonon is emitted by an electron going from the emitter to the first well, as well as from the first well to the second.

Above 21 T, the LO_0 peak moves linearly with B. An extrapolation of the LO_0 data points for $B>21$ T to $B=0$ joins the main peak at 260 mV. So, for $B>21$ T, this peak is in fact the elastic $k_\perp$-nonconserving peak ($\Delta n_T=2$) that intersects the $k_\perp$-conserving ($\Delta n_T=0$) inelastic peak at 21 T. The Landau index changes by 1 for each well crossed in this process. The elastic peak not only moves with B, but becomes much better resolved, dominating over the inelastic peak in high magnetic fields.

These two processes occur at the same bias when the field satisfies the condition $\hbar\omega_{LO}=eB/m^*$. Using the GaAs LO-phonon energy (=36 meV) and the field at the intersection point (21.2 T), we find that electrons in the quantum wells have an effective mass of $0.068m_0$. This value is in agreement with the effective mass in GaAs, supporting the above interpretation.

At values of V_b and B satisfying Eq. (5), $k_\perp$-nonconserving ($\Delta n>0$) inelastic tunneling occurs. In this case, going into either well, an electron emits a LO phonon and tunnels into a Landau level of *higher* index than it left. For each value of Δn_T, there is a valley-current peak whose bias-voltage position increases linearly with B. Since Δn_T can be 1,2, . . . , the slope $(\Delta V_b/\Delta B)^{-1}$ of a $\Delta n_T=1$ peak will be twice as large as that of a $\Delta n_T=2$. In addition, as $B\rightarrow 0$, these features should merge with the LO_0 peak, in contrast to the behavior of elastic valley-current peaks that extrapolate to V_p as $B\rightarrow 0$.

Figure 4 shows the LO_1, LO_2, LO_4, and LO_6 positions emerging from the LO_0 peak and increasing linearly with B. The LO_1 slope is twice the LO_2 slope, 4 times that of LO_3, and 6 times the slope of LO_6. The LO_1 slope corresponds to a $k_\perp$-nonconserving change of one Landau level across the structure ($\Delta n_T=1$). Similarly the higher-bias features are associated with $\Delta n_T=2$, 4, and 6. The observed even values of Δn_T correspond to equal voltage drops across each well, while odd values correspond to different drops across each well.

In this description of magnetotunneling, there are special biases and fields at which both elastic and inelastic processes occur. At these points, there should be relative maxima in the valley feature amplitudes (not related to any B dependence of the phonon-emission probability). We do observe relative maxima in the LO_i peaks. However, a quantitative interpretation of these results is not yet conclusive.

In summary, magnetotunneling experiments on a DQWTS permitted a detailed analysis of selection-rule-violating tunneling mechanism. The B-induced suppression of the $k_\perp$-nonconserving part of I_p puts an upper limit on the amount of resonant (coherent) tunneling current). The various valley-current replica features were shown to arise from elastic or inelastic tunneling processes. Our analysis enabled us to determine *directly* the effective mass of tunneling electrons.

The authors are grateful for the hospitality of the Service National des Champs Intenses, Grenoble, and the Francis Bitter National Magnet Laboratory (FBNML), at the Massachusetts Institute of Technology, especially for the hybrid magnet time at the FBNML, without which key results could not have been obtained. We would like to thank M. McLennan, for his model calculations, and W. R. Frensley, T. W. Hickmott, D. Thomas, and A. Zaslavsky for useful conversations. This work was supported in part by the U.S. National Science Foundation (NSF) under Grant No. DMR87-19634 and NSF–Centre National de la Recherche Scientifique (CNRS, France) collaboration under Grant No. INT-88-15314.

*Also at Francis Bitter National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139. Present address: Alcatel-Espace, 31037 Toulouse, France.

[1]E. Wolak, K. L. Lear, P. M. Pitner, E. S. Hellman, B. G. Park, T. Weil, J. S. Harris, Jr,, and D. Thomas, Appl. Phys. Lett. **53**, 201 (1988).

[2]E. E. Mendez, L. Esaki, and W. I. Wang, Phys. Rev. B **33**, 2893 (1986).

[3]M. L. Leadbeater, E. S. Alves, L. Eaves, M. Henini, O. H. Hughes, A. C. Celeste, and J. C. Portal, in Proceedings of the 5th Conference on Superlattices, Microstructures and Microdevices, Trieste, 1988, [Superlatt. Microstruct. (to be published)].

[4]V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **36**, 7635 (1987).

[5]H. Bando, T. Nakagawa, H. Tokumoto, K. Ohta, and K. Kajimura, in Proceedings of the 18th Conference on Low Temperature Physics, Kyoto, 1987 [Jpn. J. Appl. Phys. **26** (Suppl. 3) (1987)].

[6]M. L. Leadbeater, E. S. Alves, L. Eaves, M. Henini, O. H. Hughes, A. Celeste, J. C. Portal, G. Hill, and M. A. Pate, Phys. Rev. B **39**, 3438 (1989).

[7]C. J. Summers, K. F. Brennan, A. Torabi, and H. M. Harris, Appl. Phys. Lett. **52**, 132 (1988).

[8]B. Ricco and M. Ya. Azbel, Phys. Rev. B **28**, 3241 (1983).

[9]T. Weil and B. Vinter, Appl. Phys. Lett. **50**, 1281 (1987).

[10]M. Jonson and A. Grincwajg, Appl. Phys. Lett. **51**, 1729 (1987).

[11]V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **35**, 9387 (1987).

[12]M. McLennan (private communication). The details of this calculation appear in M. Cahay, M. McLennan, S. Datta, and M. S. Lundstrom, Appl. Phys. Lett. **50**, 612 (1987).

[13]E. E. Mendez, E. Calleja, and W. I. Wang, Appl. Phys. Lett. **53**, 977 (1988).

[14]In contrast to what is observed for $k_{ez}=0$ electrons when BIJ, as shown by S. Ben Amor, K. P. Martin, J. J. L. Rascol, R. J. Higgins, A. Torabi, H. M. Harris, and C. J. Summers, Appl. Phys. Lett. **53**, 2540 (1988).

[15]S. Ben Amor, K. P. Martin, J. J. L. Rascol, R. J. Higgins, R. C. Potter, A. A. Lakhanini, and H. Hier, Appl. Phys. Lett. **54**, 1908 (1989).

[16]D. Thomas (private communication).

[17]C. I. Huang, M. J. Paulus, C. A. Bozada, S. C. Dudley, K. R. Evans, C. E. Stutz, R. L. Jones, and M. E. Cheney, Appl. Phys. Lett. **51**, 121 (1987).

Magnetotransport studies of charge accumulation in an AlInAs/GaInAs tunneling structure

S. Ben Amor,[a] K. P. Martin, J. J. L. Rascol, and R. J. Higgins
School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

R. C. Potter, A. A. Lakhani, and H. Hier
Allied Signal Aerospace Company, Aerospace Technology Center, 9140 Old Annapolis Road, Columbia, Maryland 21045

(Received 16 December 1988; accepted for publication 9 March 1989)

We report a study of the current-voltage characteristics of a double barrier, lattice matched, quantum well tunneling structure in a quantizing magnetic field (**B**‖**J**). Experiments were conducted at fields up to 23 T at 1.5 K. The heterostructure investigated had 400 Å spacer layers in the emitter and collector, a barrier width of 72 Å, and a 43-Å-wide quantum well. This structure showed one negative differential resistance region with a peak-to-valley ratio of 23 at 4.2 K. We observed magnetoquantum oscillations, periodic in $1/B$, associated with tunneling from a quantized state in the emitter. The overall magnetoconductance dramatically changed with applied bias. We associated these variations with a field-induced increase of the impedance of the undoped spacer layers. The frequency of these oscillations increased linearly with applied bias. A discontinuity in this dependence is observed around the peak bias voltage which is the direct result of the dynamical storage and release of charge in the well.

AlInAs/GaInAs quantum well tunneling structures (QWTS) have recently raised considerable interest due to their high room-temperature peak-to-valley ratio and large current densities[1,2] compared to GaAs/GaAlAs QWTS.[3,4] Magnetotunneling is a powerful technique for the investigation of QWTS[5–8]; the effect of a quantizing field (**B**‖**J**) was used to investigate such fundamental issues as the occurrence of either sequential or resonant tunneling and the intrinsic character of the bistability.[9] In this letter, we report a magnetotunneling study on a lattice-matched AlInAs/GaInAs QWTS. After describing the samples and experimental conditions, we discuss the effect of magnetic fields on the voltage distribution in the sample. The observation of magnetoquantum oscillations allows us to study the charge accumulation in the emitter and the voltage drops in different regions of the QWTS. We also show evidence of the dynamical storage and release of charge in the well at peak voltage bias (V_p). These results are compared with work on GaAs/AlGaAs QWTS with narrow spacer layer[7]; we show that the differences originate from the different doping profile and voltage division across the samples.

The tunneling structure was grown by molecular beam epitaxy. It consisted of an n^+-InP substrate, a 0.5 μm n^+ (Si doped 2×10^{18} cm^{-3}) GaInAs layer, a 400 Å undoped GaInAs spacer layer (with a residual impurity level in the upper 10^{15} cm^{-3} range), a 72 Å undoped AlInAs barrier, a 43 Å undoped GaInAs well, a 72 Å undoped AlInAs barrier, a 400 Å undoped GaInAs spacer layer, and finally a 0.5 μm n^+-GaInAs layer (2×10^{18} cm^{-3}). This structure was fabricated into devices with mesas of sizes ranging from 25 to 400 μm^2 (further details about growth and fabrication may be found in Ref. 2). The current-voltage (I-V) characteristic of these devices showed a peak-to-valley ratio of 6 at room temperature and 23 at 4.2 K, with small variations from mesa to mesa. The measurements were performed at 1.5 and 4.2 K in magnetic fields up to 23 T. To minimize the effect of the measurement circuit on the stability of the device, the resistance of the voltage source and probe leads was kept less than 10 Ω. Nevertheless only the smallest (and highest impedance) structures (25 and 50 μm^2) were intrinsically stable. All data presented here were obtained on structures which were stable throughout the whole range of bias voltage (V_b) and magnetic field (B), allowing us to investigate the effect of magnetic field in the (sometimes inaccessible) negative differential resistance region.

Figure 1 shows that the overall shape of the I-V characteristic is not strongly affected by the presence of a magnetic field. There is a slight variation in the peak position ($V_p = 0.675$ V at $B = 0$) and peak current (I_p), and at high enough B (above 15 T) additional structures can be seen in the valley current. The valley current features are similar to those observed in GaAs/AlGaAs based QWTS and are related to off-resonance phonon emission.[10] A detailed study of these processes in our samples will be published elsewhere. At fixed V_b we observe strong magnetoquantum oscillations of the tunneling current (Fig. 2) at all biases larger than threshold (around 0.3 V). We also note that depending on the applied bias, the B dependence of the current (at fixed V_b) shows either negative, positive, or flat magnetoconductance.

Figure 2 shows that the overall background about which the B-induced oscillations occur either decreases [$V_b < V_p$, Fig. 2(a)], increases [V_b slightly higher than V_p Fig. 2(b)], or remains flat [valley current region Fig. 2(c)] depending upon where V_b is on the I-V curve (Fig. 1). This is represented in more details in the inset to Fig. 1 where the difference

[a] Visiting Scientist at Francis Bitter National Magnet Laboratory, M. I. T., Cambridge, MA 02138.

 0003-6951/89/191908-03$01.00

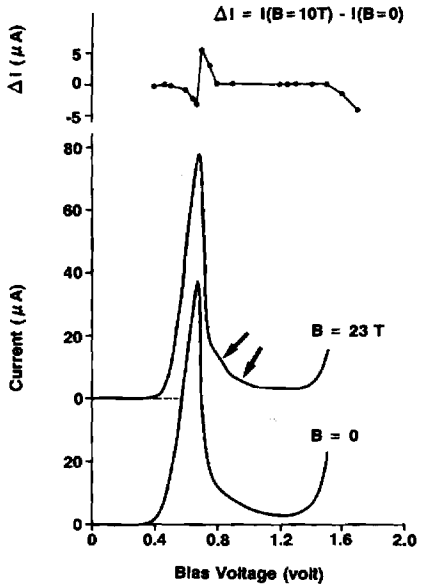


FIG. 1. I-V at $B = 0$ and 23 T, the inset shows the variation of current at $B = 10$ T relative to $B = 0$ T as a function of bias voltage.

between the I-V curve at 10 T and 0 field is plotted. At fixed bias B decreases the current for all values of $V_b < V_p$, increases it for $V_b > V_p$ (but before the valley current region), leaves it unaffected in the valley current region, and decreases the current for the thermoionic regime beyond the valley current. This behavior can be explained by a B-induced increase in the resistance of the undoped region away from the quantum well and barriers (such magnetoresistance effect should be small but positive in this B configuration and doping range[11]). The magnetoresistance effect slightly lowers the effective bias across the quantum well and barriers. The resulting change in current in the various por-

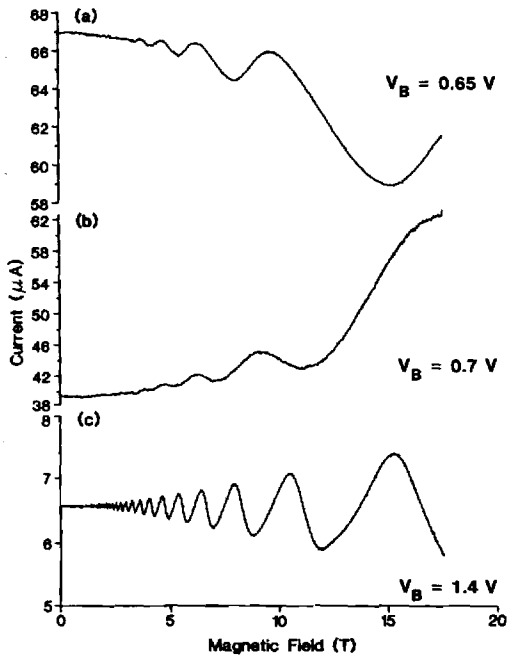


FIG. 2. I-B at different fixed bias voltages.

tions of the I-V can be visualized by introducing approximatively a 5 mV (at 10 T) decrease in bias voltage in Fig. 1 (this decrease is less than 1% of V_p and will be neglected in the rest of this study).

The magnetotunneling oscillations are periodic in $1/B$ and their frequency [defined as $\delta(1/B)^{-1} = B_f$] linearly increases with bias (represented in Fig. 3 for reverse bias) at all voltages above the threshold (0.28 V). B_f increases from 12 T (i.e., 5.8×10^{11} cm^{-2}) to 33 T (i.e., 1.6×10^{12} cm^{-2}) as the bias is swept from 0.28 to 1.3 V. At some biases up to 20 oscillations are directly resolved with the first developing at fields as low as 1.5 T. A 1.5 T onset field corresponds to an energy resolution of less than 4 meV, comparable to the broadening of the quantized level in the well for one monolayer difference in width. This observation, along with the carrier densities measured, and the *monotonic* increase of B_f with bias shows that the oscillations can only originate in the quasi-two-dimensional accumulation layer. These frequencies are higher than those observed previously[7]; the difference is likely due to the high Fermi level (125 meV) in the emitter and to a lower effective mass in our structures.

For a nonscattered electron, momentum (transverse to the tunneling direction) is conserved; in a parallel field, the momentum will be quantized in Landau levels. The energy selection rule is

$$E_a + (n + 1/2)\hbar\omega_{ca} = E_q + (p + 1/2)\hbar\omega_{cq},$$

where E_a and $\hbar\omega_{ca}$ are respectively the z-quantized energy and cyclotron energy in the accumulation layer. Similarly E_q and $\hbar\omega_{cq}$ are for the square well and n and p are the respective Landau level indices. As we have seen, the position of the resonance peak does not change with the field and therefore the two Landau ladders have equal spacing [$\hbar\omega_{ca} = \hbar\omega_{cq}$, implying that nonparabolicity effects cancel each other as the Fermi (125 meV) and quantized (in the well $E_q = 150$ meV) energies are not very different]. The conservation of momentum now corresponds to conservation of Landau level indices $n = p$. In the present case we attribute the oscillations to Landau levels passing through the Fermi level of the emitter. At $V_b > V_p$, the electrons tunnel nonresonantly but the oscillations remain linked to the same process.

The frequency dependence of Fig. 3 corresponds to an increase of the quasi-two-dimensional charge in the emitter accumulation layer. Slightly below V_p (at 400 mV), the periodicity saturates at 15 T (i.e., 7.3×10^{11} cm^{-2}) and remains constant up to V_p. At V_p the frequency abruptly increases (3.6 T jump) and again shows the same linear behavior as at low biases.

This is the direct result of *charge storage and release* in the well. Upon resonance, charge builds up in the well instead of the emitter accumulation layer and the frequency saturates. At biases greater than the peak, the charge dynamically stored in the well abruptly decreases and the related additional potential drop between the well and collector disappears. The effective bias voltage across the emitter accumulation layer and barrier is hence suddenly increased. This simple picture explains both the saturation and the jump in frequencies. From a phenomenological approach based on

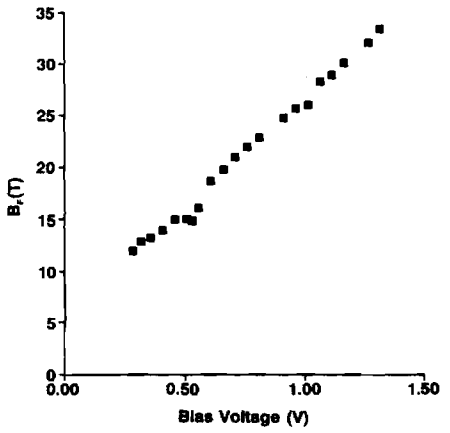


FIG. 3. Frequency of the magneto-oscillations (B_f) vs bias voltage.

Refs. 6 and 7, this jump in frequencies was estimated to result from a $2–3\times10^{11}$ cm^{-2} variation of charge in the well. This appears consistent with other estimates ranging from 1 to 4.5×10^{11} cm^{-2}.[6,7,9]

It is worth noting that in our case the frequency increases 21 T/V which is to be compared with 100 T/V obtained in Ref. 7. This discrepancy is accounted for by the existence of 400-Å-thick spacer layers in our samples compared with a 187 Å total width of the well and barriers; only a small portion of the applied bias produces a voltage drop across the well, hence the smaller slope in Fig. 3. Also, we did not observe any oscillations of the tunneling current before the current threshold [i.e., nonresonant, three-dimensional (3D) oscillations]. Any low bias 3D oscillations will reflect the high emitter doping and should have a frequency around 45 T. In the present case the frequency dependence of Fig. 3 extrapolates to 6 T at 0 bias (and not to 0 as in Ref. 7), clearly inferior to the estimated 3D frequency. The high confinement barrier (500 meV) in the present system does not allow any low bias nonresonant tunneling oscillations since the current decreases exponentially with barrier height. The finite frequency at zero bias shows that a confined emitter state exists at low (or zero) bias in these devices, likely related to the existence of a wide spacer layer.

Self-consistent calculations of the band structure in GaAs-based structures with wide spacers show the formation of a step in the conduction band in the emitter region before the accumulation layer.[12] Under bias, this step forms part of the confining potential (along with the AlInAs barrier) that quantizes the emitter energy. We believe that the formation of this potential step accounts for the observed frequency offset at $V_b = 0$. It is worth noting that Thomas *et al.*[12] report a similar observation in their study of wide spacer GaAs/AlGaAs QWTS devices.

A fan diagram of Landau levels obtained from the extrema of the magneto-oscillations permitted us to index each peak. We note that the Landau levels are linear over a wide range of B (small nonparabolicity), but with a discontinuity occurring precisely at the peak bias (this being fully consistent with the above interpretations). If the fan diagram for forward biases ($V_p = 675$ mV) is extrapolated to $B = 0$, we find an anomalous intercept of -400 mV. The reverse bias ($V_p = 515$ mV) fan diagram also shows a -300 meV intercept at zero B. The lowest V_p suggests an unintentional doping in the spacer layer on the substrate side of the QWTS (due to migration of Si).[6] The anomalous negative intercepts could be due to the larger spacer width and unintentional doping or to some non-ohmic contribution (presumably at the contacts).

Magnetotunneling measurements in the longitudinal configuration in fields up to 23 T at 1.5 K were performed on a single well, double barrier AlInAs/GaInAs QWTS. Magnetoquantum oscillations, which were attributed to tunneling from a z-quantized state in the quasi-two-dimensional emitter, were observed. The bias dependence of overall magnetoconductivity was shown to be related to a small magnetoresistance effect in the wide spacer layers. The abrupt change in the linear slope of B_f with applied bias around the peak bias was shown to be a direct result of charge dynamically accumulating (emptying) in the well.

Some of the authors (SBA, KPM, and RJH) thank B. L. Brandt and L. G. Rubin for the hospitality of the Francis Bitter National Magnet Lab. The authors are grateful to D.Beyea for device processing and E. Hempfling for device characterization. This work was supported in part by the National Science Foundation through grant DMR87-19634.

[1] F. Capasso, S. Sen, A. Y. Cho, and D. L. Sivco, Appl. Phys. Lett. **53**, 1056 (1988).
[2] R. C. Potter, A. A. Lakhani, D. Beyea, H. Hier, E. Hempfling, and A. Fathimulla, Appl. Phys. Lett. **52**, 2163 (1988).
[3] T. Inata, S. Muto, Y. Nakata, T. Fujii, H. Oshini, and S. Hiyamizu, Jpn. J. Appl. Phys. **25**, L983 (1986).
[4] R. C. Potter, A. A. Lakhani, D. Beyea, and Harry Hier, J. Appl. Phys. **63**, 5875 (1988).
[5] E. E. Mendez, L. Esaki, and W. Wang, Phys. Rev. B **33**, 2893 (1985).
[6] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **35**, 9387 (1987).
[7] L. Eaves, G. A. Toombs, F. W. Sheard, C. A. Payling, M. L. Leadbeater, E. S. Alves, T. J. Foster, P. E. Simmonds, M. Henini, O. H. Hughes, J. C. Portal, G. Hill, and M. A. Pate, Appl. Phys. Lett. **52**, 212 (1988).
[8] S. Ben Amor, K. P. Martin, J. J. L. Rascol, R. J. Higgins, A. Torabi, H. M. Harris, and C. J. Summers, Appl. Phys. Lett. **53**, 2540 (1988).
[9] A. Zaslavsky, V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Appl. Phys. Lett. **53**, 1408 (1988); M. L. Leadbeater, E. S. Alves, L. Eaves, M. Heinini, O. H. Hughes, F. W. Sheard, and G. A. Toombs, Semicond. Sci. Technol. **3**, 1060 (1988).
[10] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. B **36**, 7635 (1987).
[11] J. A. Chrobosheck, L. Eaves, P. S. S. Guimares, P. C. Main, I. P. Roche, H. Mitter, J. C. Portal, P. N. Butcher, M. Ketkar, and S. Summerfield, in *Proceedings of 17th International Conference on the Physics of Semiconductors, San Francisco, CA, 1984* (Springer, Berlin, Heidelberg, New York, 1984), p. 696.
[12] D. Thomas, F. Chevoir, P. Bois, E. Barbier, Y. Guldner, and J. P. Vieren, Proceedings of the 5th Conference on Superlattices, Microstructures and Microdevices, Trieste, 1988, to be published in Superlattices and Microstructures.

Transverse magnetic field dependence of the current-voltage characteristics of double-barrier quantum well tunneling structures

S. Ben Amor,[a] K. P. Martin, J. J. L. Rascol, and R. J. Higgins
School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

A. Torabi, H. M. Harris, and C. J. Summers
Georgia Tech Research Institute and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 15 August 1988; accepted for publication 17 October 1988)

We report the effects of a transverse magnetic field (J⊥B) on the conductivity of quantum well tunneling structures based on AlGaAs/GaAs/AlGaAs quantum wells. The current-voltage characteristics in the positive differential resistance regime show negative magnetoconductance for all values of B. The peak bias voltage increases monotonically with increasing B. For $B<6$ T there is a decrease in the peak tunneling current, but then it increases for $B>6$ T. The data also show dramatic magnetic field induced changes in the negative differential resistance (NDR) features. The behavior of the NDR changes from sharp hysteretic bistable-like transitions to astable NDR transitions. Both the valley current and its bias voltage position increase with increasing magnetic field. This behavior is described by a simple model that includes magnetic field effects across the barriers.

Quantum well tunneling structures (QWTSs) have attracted considerable interest since they offer the experimental opportunity to investigate such fundamental quantum effects as tunneling and wave function phase coherence. QWTSs have also shown promise in device application as high-speed switches and high-frequency oscillators.[1] Magnetotunneling studies have proved to be a powerful tool to study the dynamics of electrons in these structures. The effect of a longitudinal magnetic field (J||B) on the tunneling current has already been used to probe the difference between sequential and resonant tunneling.[2] In this letter the effect of a magnetic field applied perpendicular to the current is investigated.

A motivation for this work is to alter the tunneling through the double-barrier structure for a specific bias voltage without using a base contact. From a classical point of view a Lorentz force is applied by B to electrons incident to the quantum well and barriers. An important consequence is to change the transverse momentum of the electrons and decrease their kinetic energy along the incident direction. A weaker effect arises from the diamagnetic shift that increases the energy of the quantum levels in the well. For the quantum well in this study (51 Å wide) the diamagnetic shift is calculated to be a few meV at 10 T. Also, since the path of the electron through the whole structure is increased, some contribution from magnetoresistance effects may be expected.

The samples used in this work were grown by molecular beam epitaxy. The QWTSs consisted of an n^+ substrate, a Si-doped 25-period (14 Å/period) $Al_xGa_{1-x}As/GaAs$ superlattice ($x=0.35$), a 0.7 μm GaAs layer doped from 5×10^{17} to 1×10^{17} cm^{-3}, a 51 Å undoped GaAs layer, the double-barrier structure, a 51 Å undoped GaAs layer, and a 0.9 μm Si-doped GaAs cap layer (graded from 1×10^{17} cm^{-3} $<n<$ 5×10^{17} cm^{-3}). The double-barrier quantum well portion was formed from 55-Å-wide AlGaAs layers separated by a 55 Å undoped GaAs layer. For this experiment 50-μm-diam dots were fabricated from this material. Further details about growth and processing techniques are described in Ref. 3.

The measurements were made at 4.2 K in magnetic fields up to 15 T. It has already been demonstrated the dramatic effect the measurement circuit has on the features in the negative differential resistance (NDR) region.[4] In this work the combined resistance of the voltage source, probe leads, and in-series current measuring resistor ($R_s=1\ \Omega$) was less than 10 Ω. In addition, some of the measurements were made with a 20 Ω resistor (mounted at the sample end of the probe) in parallel with the QWTS and R_s. This permitted stray capacitive or inductive effects of the external circuit to be reduced.

Current-voltage (I-V) curves on a representative sample labeled 103 at different magnetic fields (B perpendicular to J) are displayed on Fig. 1. Results with the same features were seen reproducibly on similar samples grown in this laboratory. Both the peak bias voltage (V_p) and valley bias voltage (V_v) increase rapidly with B. The peak to valley ratio decreases with increasing magnetic field until the NDR regime is washed out at 13 T. At a fixed bias, away from the NDR region, the tunneling current decreases monotonically with increasing magnetic field whereas the peak current (I_p) goes through a minimum at 6 T and increases for higher values of B. Another feature seen in the data is the gradual smearing out of the I-V peak as the field is increased. The B-induced increase in V_p and smearing out of the I-V peak has been reported for InGaAs/InP QWTS.[5] At zero field, the jump in bias voltage across the NDR transition is narrow and shows a hysteretic behavior. Under magnetic field, the NDR transition becomes wider and shows astable-like behavior.[6] In the rest of this letter, we will focus on the effect of magnetic field outside of the NDR regime.

[a] Visiting Scientist, Francis Bitter National Magnet Laboratory.

 0003-6951/88/512540-03$01.00

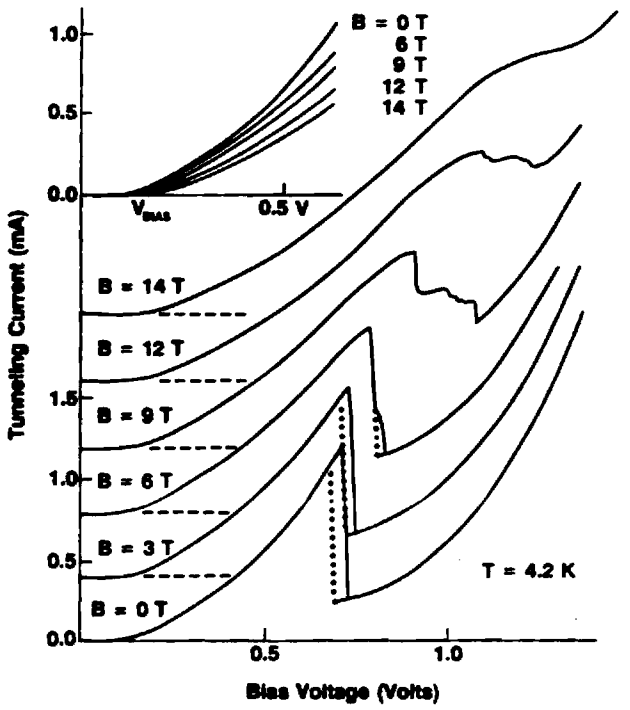


FIG. 1. Current-voltage curves of sample 103 at 4.2 K in magnetic fields up to 14 T. The zeros have been offset to separate the I-V curves taken at successively larger values of B. The large dotted lines denote the hysteretic portion when sweeping down in bias voltage. The inset (upper left) plots the I-V's for $V_{BIAS} < V_P$ at different B to illustrate the decrease in tunneling current at fixed bias voltage.

Figure 2 shows the magnetic field dependence of the peak bias voltage and current up to 13 T. V_P exhibits a quadratic increase with field, whereas there is an initial decrease of I_P, dropping by 15% at ~6 T and then increasing ~10% over the zero field value at 13 T.

The observed magnetic field dependence of V_P and I_P in our samples occurs because of changes in the tunneling selection rules and transmission probability through the barrier. I-V measurements on these samples for B parallel to J did not show strong magnetoquantum oscillations, indicating that the Landau levels in the emitter are too broad to significantly affect the density of states. Thus in the picture of sequential tunneling, the magnetic field mostly affects the selection rules for emitter to well tunneling. Our interpretation begins by focusing on the change in peak bias voltage using a simple semiclassical approach.

If the magnetic field is applied transversely along the y direction (parallel to the interface), an electron tunneling in the z direction experiences a potential increase[7,8]:

$$V(z) = e^2B^2z^2/(m) + heBk_yz/(2\pi m)\,. \qquad (1)$$

As in the zero field case, the momentum parallel to the barrier must be conserved during tunneling; hence, electrons in the emitter with momentum k_y can only tunnel to $k_y + 2\pi eBz/h$ (where z is the distance from the emitter side of the barrier to the middle of the quantum well). As the motion is not affected along the x axis, electrons tunnel to states with the same k_x in the well. In other words, the parabolic dispersion relations of the transverse momentum of the well states are shifted along the k_y axis with respect to the emitter states (Fig. 3) so an additional bias is needed for them to overlap. The first term in Eq. (1) predicts that the shift in V_P should increase quadratically with field as observed in Fig. 2. Figure 3 also shows that as B increases, the overlap between emitter and well parabolas remains over a larger bias range until the NDR is smeared out. These two major features of the data are thus readily explained by this simple framework.

The round off of the current peak can also be described by a state-space argument using this picture. The number of impinging electrons with a given k_y is proportional to $(k_f^2 - k_y^2)^{1/2}$, and electrons with different k_y do not undergo identical shifts in the effective potential [second term

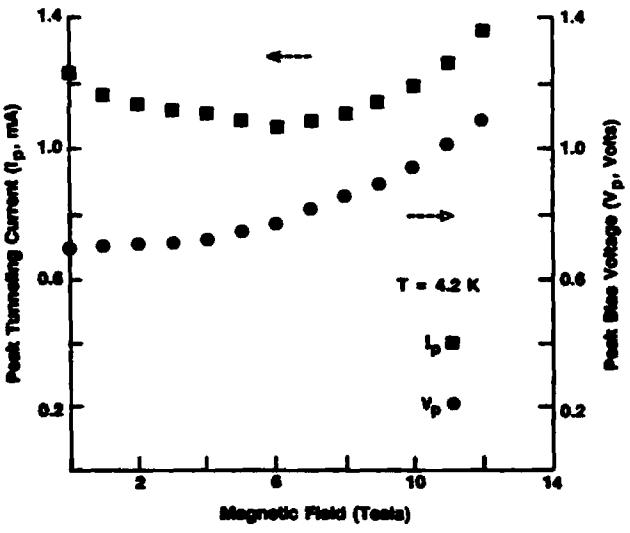


FIG. 2. Value of peak bias voltage V_P (circles) and peak tunneling current I_P (squares) of sample 103 at 4.2 K for $0 < B < 13$ T. It should be pointed out that V_P and I_P from I-V's with reversed bias show the same behavior.

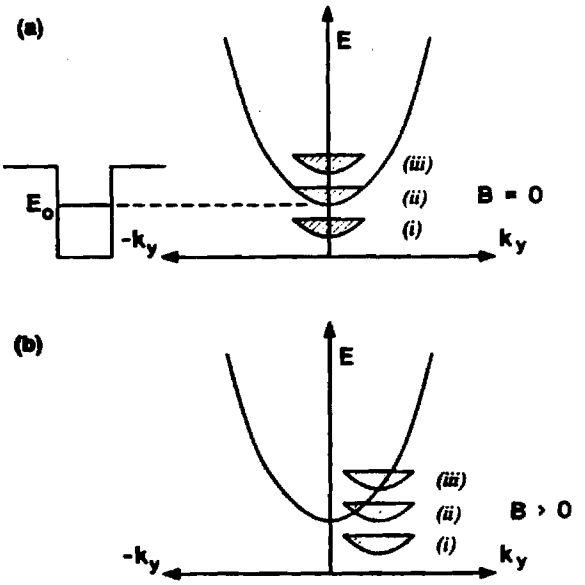


FIG. 3. Schematically shows the overlap of the occupied states in the emitter (hatched areas) and states in the quantum well (parabola) for (a) $B = 0$, and (b) $B > 0$, where (i) $V < V_P$, (ii) $V = V_P$, and (iii) $V > V_P$.

in Eq. (1)]. This k_y-dependent redistribution of electron energies results in the broadening in the tunneling current near V_P.

The experimentally observed B dependence of the tunneling current can also be described by Eq. (1) and the mechanism shown in Fig. 3. At a fixed bias (Fig. 1), the current monotonically decreases with increasing B. This can be accounted for by the perturbational approach of Eaves *et al.*,[9] where the tunneling current through a barrier (normalized to $B=0$) has been shown to decrease as $\exp(-aB^2)$, where a depends strongly on the barrier thickness and more weakly on the kinetic energy. Using the detailed expression in Ref. 9 a 5% decrease in current is predicted at 10 T. This, however, is smaller than the observed effect; we experimentally measure a current decrease (for example, at $V_{bias}=0.4$ V) of about 30 % at this field.

Several factors affect the size of the *peak* tunneling current I_P in the presence of magnetic field. V_P has already been shown to increase with B, so as the peak is shifted to higher biases, the effective transparency of the emitter barrier increases (a rectangular barrier becomes triangular) and I_P is expected to become larger. But, in general, within the picture of sequential tunneling the current depends not only on the transmission coefficient but also on the supply function of available electrons.[10] As can be deduced from Fig. 3 the overlap of emitter and well parabolas is maximum at $B=0$ so the shift in the well parabola along the k_y axis for $B>0$ decreases the overlap. As a result of $B>0$ the supply function contributes to a decrease in the tunneling current. The global effect of these two competing factors would explain the initial decrease of I_P (decrease of parabola overlap), which goes through a minimum at 6 T and increases above that field (more transparent emitter barrier) (Fig. 2).

Although the valley bias voltage moves to higher biases (Fig. 1), the valley current essentially remains on the $B=0$ I-V curve. This is expected as the existence of *any* valley current is due to the relaxation of the tunneling selection rules and is not strongly affected by the magnetic field.

The question of magnetoresistance effects in the doped emitter and collector regions has also been considered. An estimate of the mobility using the doping profile[11] in these regions indicates that a geometric-magnetoresistance effect (the increased path length from the Hall angle) contributes less than 10% to the overall conductivity at 10 T.

In conclusion, I-V measurements in transverse magnetic fields up to 15 T at 4.2 K on a double-barrier quantum well tunneling structure show that the peak and valley bias voltages increase monotonically with B. At fixed bias voltage, the tunneling current decreases with increasing B. The peak tunneling current initially decreases with increasing B, going through a minimum at 6 T. The tunneling peak is broadened until the NDR transition is washed out at $B=13$ T.

The shift towards higher voltages of V_P is accounted for by the field-induced redistribution in **k** space of the electron's kinetic energy. The behavior of I_P is the result of two competing effects, smaller overlap of emitter and well states as opposed to an increased transparency of the tunneling barrier. Although these arguments qualitatively describe all the observed features, a more sophisticated approach is necessary to adequately explain the amplitude of these effects.

Some of the authors (S. B. M., K. P. M., and R. J. H.) would like to thank B. L. Brandt and L. G. Rubin for their help and the hospitality of the Francis Bitter National Magnet Lab (at MIT). We also thank K. F. Brennan for useful discussions. This work was supported in part by the National Science Foundation under grant DMR87-19634.

[1] T. C. L. G. Sollner, W. D. Goodhue, P. E. Tannenwald, C. D. Parker, and D. D. Peck, Appl. Phys. Lett. **43**, 588 (1983).
[2] L. Eaves, G. A. Toombs, F. W. Sheard, C. A. Payling, M. L. Leadbeater, E. S. Alves, T. J. Foster, P. E. Simmonds, M. Henini, O. H. Hughes, J. C. Portal, G. Hill, and M. A Pate, Appl. Phys. Lett. **52**, 212 (1988). See also E. E. Mendez, L. Esaki, and W. I Wang, Phys. Rev. B **33**, 2893 (1986).
[3] C. J. Summers, K. F. Brennan, A. Torabi, and H. M. Harris, Appl. Phys. Lett. **52**, 132 (1988).
[4] E. S. Hellman, K. L. Lear, and J. S. Harris, Jr., Appl. Phys. Lett. **53**, 201 (1988). See also Jeff F. Young, B. M. Wood, H. C. Liu, M. Buchanan, D. Landheer, A. J. Spring Thorpe, and P. Mandeville, Appl. Phys. Lett. **52**, 1398 (1988).
[5] M. L. Leadbeater, L. Eaves, P. A. Claxton, G. Hill, M. A. Pate, and P. E. Simmons, The 5th Int. Conf. on Hot Carriers in Semiconductors, Boston, MA, 1987.
[6] We point out that the term astable is used qualitatively and should be employed carefully as the exact stabilization conditions of those structures are not fully understood.
[7] L. Eaves, K. W. H. Stevens, and F. W. Sheard, The Physics and Fabrication of Microstructures, Springer Series in Physics **13**, 343 (1986).
[8] R. A. Davies, D. J. Newson, T. G. Powell, M. J. Kelly, and H. W. Myron, Semicond. Sci. Technol. **2**, 61 (1987).
[9] L. Eaves, D. C. Taylor, J. C. Portal, and L. Dmowski, in *Proceedings of the International Winter School, Mauterndorf* (Springer, Berlin, 1986), pp. 96–105.
[10] C. B. Duke, *Tunneling in Solids* (Academic, New York, 1969), Chap. 5.
[11] G. E. Stillman and C. M. Wolfe, Thin Solid Films **31**, 69 (1976).

Phonon-assisted tunneling in persistent-photocurrent decay

L. X. He
Department of Physics, University of Oregon, Eugene, Oregon 97403

K. P. Martin* and R. J. Higgins
School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 27 June 1988)

While in bulk $Al_xGa_{1-x}As$ there is some agreement that persistent photoconductivity (PPC) is related to the photoionization of deep levels that have a recombination barrier, a diverse collection of models has been proposed to explain the interplay between macroscopic (band bending and tunneling) and microscopic (deep levels) contributions to PPC and related phenomena in $GaAs/Al_xGa_{1-x}As$ heterostructures. One of these related effects is transient photoconductivity (TPC) whereby the PPC-enhanced carrier density (or conductivity) undergoes a long-term decay. In this work a model based on phonon-assisted tunneling through a realistic barrier is developed for TPC, and predictions for the apparent capture energies and lifetime prefactors are presented. The model self-consistently accounts for the changes in the conduction-band profile and tunneling barrier due to the transfer of charge during this process. By using the derivative of conductivity or carrier number density with respect to the logarithm of time during decay, e.g., $dN/d(\ln t)$, phonon-assisted tunneling may be identified from other PPC-associated decay mechanisms in $GaAs/Al_xGa_{1-x}As$ heterostructures. This method of analysis is applied to experimental data showing TPC decay of carrier density and conductivity. The *DX*-center capture energies obtained from phonon-assisted-tunneling data in this work agree with results from other direct-capture studies.

I. INTRODUCTION

Persistent photoconductivity (PPC) is well known for many semiconductor materials and configurations.[1] It is important to understand PPC and related effects in modulation-doped materials and high-electron-mobility transistors (HEMT's) formed from $GaAs/Al_xGa_{1-x}As$ heterostructures at $T < 180$ K where HEMT's display optimal electronic properties. In bulk materials, PPC is generally understood to occur by the photoionization of deep levels that have a recombination barrier. However in $GaAs/Al_xGa_{1-x}As$-based heterostructures and devices a variety of models have been proposed to explain the interplay between macroscopic[2,3] (band bending) and microscopic[4,5] (deep levels) mechanisms for PPC. Microscopic-barrier models are based on postulated atomic-scale barriers that suppress recombination of free carriers. The carriers are photoexcited from the impurity center, a deep-donor complex (predominantly identified as the *DX* center in Si-doped $Al_xGa_{1-x}As$) with optical thresholds smaller than the band gap of the material. The origin of these deep donors has not been fully clarified but they are usually described as impurity-atom-plus-defect complexes with large lattice relaxations. The earliest model[5] for *DX* centers in doped $Al_xGa_{1-x}As$ is a donor-vacancy (As) complex, which requires an As vacancy number density as high as 1×10^{18} cm^{-3}. Recently a donor-self-generated-vacancy model has been proposed to avoid this high density of As vacancies.[6]

The decay of PPC-enhanced carrier density and mobility has been experimentally observed by a number of workers.[3,7-9] Transient photoconductivity (TPC) occurs on time scales as long as 10^5 sec and at temperatures as low as 4.2 K. Tunneling[3] provides an important mechanism for electrons to move between the GaAs and $Al_xGa_{1-x}As$ layers despite the macroscopic barrier presented by the conduction band (Fig. 1). In this work a TPC-decay model based on tunneling between the two-dimensional electron gas (2D EG) and states in the $Al_xGa_{1-x}As$-layer conduction band is presented. The transfer of electrons is accompanied by the corresponding self-consistent changes in the conduction-band profile and tunneling barrier. The model incorporates a realistic representation of the heterojunction barrier and band discontinuity into a Wentzel-Kramers-Brillouin (WKB) calculation of the tunneling probability. The lifetime of tunneling-assisted capture is determined by the tunneling probability and the capture cross section of the traps. When kT is much smaller than the height of the tunneling barrier, a WKB calculation shows that the tunneling probability only depends on the tunneling-barrier shape. At high enough temperatures, phonon-assisted tunneling becomes important.

The result of this model is a prediction of apparent capture energies and lifetime prefactors of the decay processes. Experimental measurements of persistent-photocurrent decay in different gated and ungated heterostructure material are shown to agree quite well with

these predictions. The data is analyzed by a method using a derivative of number density (or conductivity) with respect to the logarithm of time. The decay features brought out by this analysis are identified with different decay processes.

II. THEORETICAL MODEL

When the 2D EG is confined in the potential well at the $Al_xGa_{1-x}As/GaAs$ interface (Fig. 1), the z component of the electron wave function is separated from the xy component, and the z-direction energy is quantized with a ground-state energy E_0 above the conduction band. TPC tunneling is in the z direction and starts from the $Al_xGa_{1-x}As/GaAs$ interface with a carrier energy of E_0 plus the 2D EG thermal energy E_{th}. The effective tunneling barrier is the conduction-band profile above E_0+E_{th} (Fig. 1).

As the tunneling electrons first move into the conduction band in the $Al_xGa_{1-x}As$ layer, the tunneling barrier is different for the remaining 2D electrons [short-dashed curve, Fig. 1(a)] because of the decreased 2D EG density (N_{2D}) and the decrease in the depletion width W_d. This

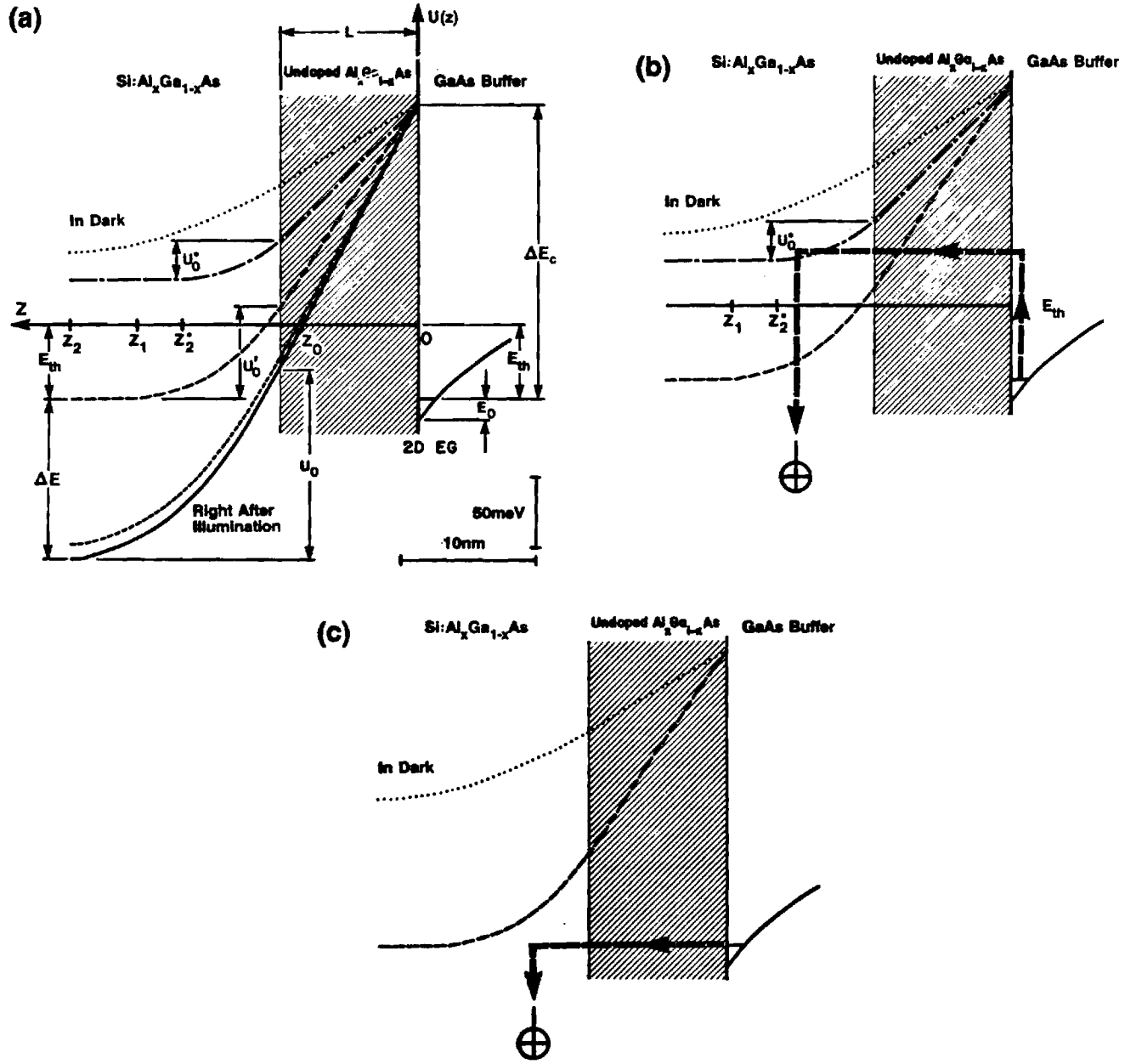


FIG. 1. Tunneling mechanisms and typical conduction-band tunneling barriers $U(z)$ (see the Appendix) for a $GaAs/Al_{0.3}Ga_{0.7}As$ heterostructure. The dotted curve is the barrier at 77 K in-dark (100 meV above E_0) (Ref. 8). The solid curve (a) is the barrier right after illumination ($W_d=15$ nm, $\Delta E=120$ meV). The long-dashed curve ($W_d=11$ nm, $\Delta E=0$) is at the end of type-A tunneling. After further decay, the barrier shown as a dot-dashed curve ($W_d=7$ nm, the conduction band is 80 meV above E_0) allows only type-B tunneling (b) and electron-impurity tunneling (c) to occur. The scales of the drawings are from the actual conditions of doping ($N_d=1\times10^{18}$ cm^{-3}) and illumination used.

larger tunneling barrier reduces the tunneling probability, and consequently the next carriers through have longer tunneling lifetimes. This process (designated as type-A tunneling) is completed when $W_d=(z_1-L)$ and the lowest value of the conduction band in the $Al_xGa_{1-x}As$ layer, E_{CB} is in equilibrium with the 2D EG. At this point the conduction-band bending appears as the long-dashed curve in Fig. 1(a).

In other circumstances, when E_{CB} moves above E_0 for further photodecay [dot-dashed curve in Fig. 1(b)] the 2D electrons will not be able to tunnel and stay in the $Al_xGa_{1-x}As$ layer unless there are unoccupied traps below E_0. This will be referred to as type-B tunneling-assisted decay.

Consider a heterostructure with a relatively small band discontinuity at a high temperature (e.g., $\Delta E_c=0.23$ eV, and $kT=0.01$ eV at 100 K) and a doping density $N_d=10^{18}$ cm^{-3}. The pure-tunneling probability (10^{-6}–10^{-10}) estimated from the WKB integral of the conduction-band profile is comparable with the thermal-activation probability $\exp(-\Delta E_c/kT)$. In this case TPC can be described as a phonon-assisted process with a tunneling probability:

$$P=A\int \exp\left[-E_{\rm th}/kT - \int (4\pi/h)[2m^*U(z)]^{1/2}dz\right]dE_{\rm th}\ , \qquad (1)$$

where m^* is the effective mass of an electron in the $Al_xGa_{1-x}As$ layer, and $E_{\rm th}$ is the thermal energy of the tunneling electron. $U(z)$ is the tunneling barrier measured from $E_0+E_{\rm th}$. The WKB integral ($\int (4\pi/h)[2m^*U(z)]^{1/2}dz$) is evaluated from $z=0$ to z_0, where z_0 is the point where the $Al_xGa_{1-x}As$ conduction band is even with $E_0+E_{\rm th}$ [Fig. 1(a)]. When the WKB integral is bigger than $E_{\rm th}/kT$, the normalization factor A is only weakly temperature dependent. Detailed expressions for $U(z)$ and z_0 are shown in the Appendix for a given $E_{\rm th}$, ΔE_c, L, and ionized donor density N_i.

Figure 2 shows the temperature-dependent tunneling lifetime τ numerically determined from Eq. (1) over a range of W_d. It is clear that at a fixed temperature, τ monotonically *increases* during TPC decay while W_d *decreases*. The lifetime can increase as much as four orders of magnitude for a factor of 2 change in W_d. It should also be noted that for type-A decay τ is weakly temperature dependent for $T<100$ K, but becomes strongly temperature dependent above 100 K because of phonon-assisted tunneling. An activation energy E_t can be determined from the slope of τ versus $1/T$ for a fixed W_d, and in the high-temperature limit E_t would approach ΔE_c.

Actually, N_i (following illumination) is a function of T because at high temperatures the deep-trap recombination rate can be comparable with the photoemission rate. Using the temperature dependence of N_i,[10] (for an Al fraction $x=0.3$) results in an apparent $E_t\sim 170$ meV at 160 K, but about 20 meV at 70 K, for $W_d=(z_1-L)$ and $N_d=1\times10^{18}$ cm^{-3}.

Although the difference in mobility between 2D electrons and carriers in the $Al_xGa_{1-x}As$ layer allows pure tunneling to be detected, the times (10^{-8}–10^{-4} sec) (Ref. 11) are too short to be observed in this experiment. When this process is completed, direct capture of electrons in the $Al_xGa_{1-x}As$ layer by nearby traps can be detected, generally with a longer, trap-dependent lifetime. In the case of shallow traps, τ is assumed to be weakly temperature dependent, and on the order of 10^{-3}–10^{-2} sec. However for deep traps such as DX centers, τ is strongly temperature dependent.

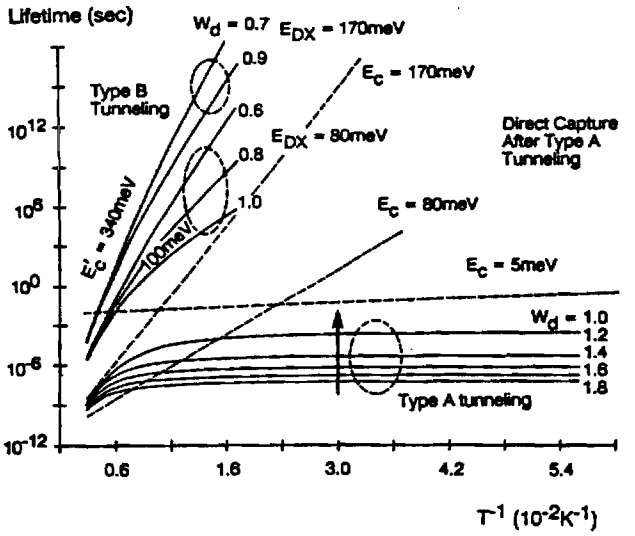


FIG. 2. The predicted TPC lifetimes. The arrow denotes the time evolution of type-A tunneling for fixed T [starting from $W_d=1.8(z_1-L)$ and ending at $W_d=1.0(z_1-L)$]. Also shown is type-B tunneling-assisted decay with DX-center capture energies of either 80 meV [decay from $W_d=1.0(z_1-L)$ to $W_d=0.6(z_1-L)$] or 170 meV [decay from $W_d=0.9(z_1-L)$ to $W_d=0.7(z_1-L)$]. For comparison, lifetimes from direct capture in the $Al_xGa_{1-x}As$ layer by shallow ($E_c=5$ meV) or deep traps ($E_c=80$ or 170 meV) after type A are represented as dashed lines.

The capture lifetime of a conduction-band electron in the $Al_xGa_{1-x}As$ layer trapped by a DX center is expressed as

$$\tau=\tau_0\exp(E_{DX}/kT)\ , \qquad (2)$$

where E_{DX} is the capture energy, and $\tau_0\sim10^{-11}$ sec.

So in the case of type-B tunneling the lifetime of a 2D electron captured by DX centers is expressed by

$$\tau=[\tau_0\exp(E_{DX}/kT)]P^{-1}\ . \qquad (3)$$

Here the tunneling probability is calculated from Eq. (1) and the Appendix, with the integral over $E_{\rm th}$ beginning from E_{CB} [dashed line, Fig. 1(b)].

Over a narrow temperature range, the phonon-assisted tunneling probability P may be expressed as $P=P_0\exp(-E_t/kT)$, and one has

$$\tau=\tau_0^*\exp[(E_{DX}+E_t)/kT] \qquad (4)$$

with $\tau_0^*=\tau_0P_0^{-1}$ and an apparent capture energy $E_c'=E_{DX}+E_t$. Notice that E_t and P are temperature

dependent. Considering that $P_0=1$ for phonon-assisted tunneling in the high-temperature limit, the value of τ_0^* should be the same as the case for direct capture by deep traps at high temperatures, i.e., $\tau_0^* \sim 10^{-11}$ sec. At lower temperatures ($T \sim 77$ K), $P_0 \ll 1$, and $\tau_0^* \gg \tau_0$.

The calculated results for type-B tunneling-assisted capture into deep traps are also plotted in Fig. 2 using DX-center capture energies $E_c=80$ and 170 meV. Note that E_c' taken from the slope of the curves in Fig. 6 is 100 meV at 77 K for $E_{DX}=80$ and 340 meV at 160 K for $E_{DX}=170$ meV.

For comparison direct capture in the $Al_xGa_{1-x}As$ layer by shallow or deep traps after type-A tunneling is plotted in Fig. 2 as dashed lines, with shallow-trap capture energies equal to 5 meV, and deep-trap capture energies chosen as 80 and 170 meV.

Another type of tunneling process that could contribute to TPC is electron-impurity tunneling[3] [Fig. 1(c)]. This mechanism suggests that 2D electrons tunnel into the $Al_xGa_{1-x}As$ depletion region and then are captured by deep traps. The trapping begins with donor traps nearest the 2D EG. While the initial 2D EG states and final trapped states have to be real, the intermediate states (e.g., in the forbidden gap) do not have to be real, but can be regarded as virtual states.[12] Though the detailed nature of the "virtual states" are not fully clarified, recapture dynamics of this tunneling would follow a decay curve with a much longer lifetime prefactor due to a small tunneling probability. For example, experiments at temperatures below 100 K have shown[3] that the lifetime prefactors are of the order of 1 sec, and the effective capture energy is estimated to be below 10 meV.

We now summarize the key results from Fig. 2 and make some predictions (excluding electron-impurity tunneling). For $T<50$ K, the observable capture within a time window of $10^{-3}-10^4$ sec is only shallow-donor related. For higher temperatures both shallow and deep traps can be involved.

Because the $Al_xGa_{1-x}As$-layer conduction band moves up with respect to E_0 during photodecay, tunneling-assisted capture can change from type-A to type-B tunneling. When all donors are photoionized [Fig. 1(a)] there is a large depletion width and low $Al_xGa_{1-x}As$ conduction-band minima so that type-A tunneling-assisted capture initially takes place. After a sufficiently long decay time, the $Al_xGa_{1-x}As$ conduction band is above E_0 but below the original in-dark conduction band [Fig. 1(b)], and then type-B tunneling-assisted capture takes place.

As an example, consider the case of $T=77$ K in Fig. 2. Direct capture by 80 meV deep centers and type-A tunneling occurs too quickly to be detected within an experimental time window of $10^{-3}-10^4$ sec of this work. Instead the first process to be observed is direct capture by shallow donors. Then, if the $Al_xGa_{1-x}As$-layer conduction band moves above E_0, type-B tunneling-assisted decay (with 80 meV deep traps) can be detected. E_c' will be about 100 meV due to the additional E_{th} at 77 K, and $\tau_0 \sim 10^{-5}$ sec (y intercept of the slope in Fig. 2), which is much longer than the DX-center capture prefactor of 10^{-11} sec. Alternatively, if the $Al_xGa_{1-x}As$ layer conduction band does not move above E_0, as in the case of a wide, doped layer with a large number of free electrons in it, direct capture to 170 meV deep traps dominates.

When the temperature is as high as 160 K, the first process to be seen within the same time window ($10^{-3}-10^4$ sec) is direct capture by shallow traps. The next process to appear is type-B tunneling-assisted capture by 170 meV deep traps. In this case, E_c' (about 340 meV) will be much large than the DX-center capture energy 170 meV, but the lifetime prefactor remains close to the value of 10^{-11} sec.

III. SAMPLES AND MEASUREMENTS

The molecular-beam-epitaxy- (MBE-) grown heterostructure samples in this work had the following layer sequence: Si-doped GaAs cap layer, doped $Al_xGa_{1-x}As$ layer, thin undoped $Al_xGa_{1-x}As$ spacer layer, thick undoped GaAs buffer layer, and semi-insulating GaAs substrate. A two-dimensional free-electron gas is formed at the interface between the undoped $Al_xGa_{1-x}As$ and GaAs buffer layers, separated by the spacer layer from the nearest donors in $Al_xGa_{1-x}As$ layer. A typical device, the so-called high-electron-mobility transistor, has a metal gate with a conductive channel width much larger than the distance between source and drain (>10:1). Other samples employ a van der Pauw (VDP) square geometry with an ohmic contact in each corner.

Samples from two different sources were used. Summarized in Table I are their principal parameters: the

TABLE I. Principal parameters of samples. Sample label, cap-layer thickness L_c, cap-layer doping density $N_{d,c}$, $Al_xGa_{1-x}As$ layer thickness L_a, $Al_xGa_{1-x}As$ layer doping density $N_{d,a}$, undoped spacer thickness L_s, gate voltage threshold (V_t) at 77 K in the dark, the net ionized donor density N_i in depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, and the percentage increase of N_i after identical brief white light illumination sufficient to saturate the PPC change. Layer thicknesses are in units of nm and doping/carrier densities are in units of 10^{17} cm^{-3}. All samples had an aluminim fraction $x=0.3$ in the doped $Al_xGa_{1-x}As$ layer.

Sample	L_c	$N_{d,c}$	L_a	$N_{d,a}$	L_s	V_t	N_i	$\Delta N_i/N_i$
B^a	50	2	100	10	5	−0.5 V	1.9	15%
C^b,D^b	20	10	30	10	10		ungated VDP	

[a]Tek: Tektronix Inc., R. Gleason and R. Koyama.
[b]Cornell: Cornell University, W. Schaff.

cap-layer thickness L_c, cap-layer doping density $N_{d,c}$, $Al_xGa_{1-x}As$ layer thickness L_a, $Al_xGa_{1-x}As$ layer doping density $N_{d,a}$, and undoped spacer thickness L_s. Table I includes the threshold gate voltage (V_t), and net ionized donor density N_i in the depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, as well as the percentage increase of N_i after saturated PPC for sample B. Sample D, an ungated VDP square, had $x=0.3$ in the Si-doped $Al_xGa_{1-x}As$ layer but $x=1.0$ in the undoped spacer layer and a 10 nm undoped $Al_xGa_{1-x}As$ layer ($x=0.3$) between the cap layer and the doped $Al_xGa_{1-x}As$ layer. Sample C has layer parameters similar to sample D but a different annealing temperature. For samples C and D (with relatively thin doped $Al_xGa_{1-x}As$ layers, $L_a=30$ nm) the in-dark band bending is like that shown in Fig. 1 and there are few free electrons in the $Al_xGa_{1-x}As$ layer. Table II shows the in-dark and following-light exposure carrier density and mobility of samples C and D at 77 K.

For sample B (a gated HEMT), the source-drain current I_{SD} was measured with a fixed source-drain voltage. For the ungated van der Pauw samples C and D, the carrier densities N_s were measured via the Hall effect. For example, [Fig. 3(a)] after illumination is removed, N_s of sample C decays with an initially fast decay, followed by a continuously decreasing decay rate. The decay is nearly a straight line when plotted as a function of $\log_{10}(t)$ for close to seven orders of magnitude in time indicating a wide range of the decay time constant [Fig. 3(b)]. Similar decay measurements of I_{SD} from sample B show this nonexponential result.

IV. ANALYSIS USING ln(t)-DERIVATIVE METHOD

In view of the possible decay mechanisms previously described detailed information about capture energies and lifetime prefactors is needed to distinguish among the different processes that contribute to TPC. In the following, we demonstrate how the temperature-dependent decay time can be analyzed using a ln(t) derivative technique (full details are found in Ref. 13). This yields a trap spectroscopy with characteristic energies and lifetime prefactors that differ so significantly for different capture mechanisms that phonon-assisted tunneling and other decay channels may be unambigously identified.

Assume a general form of multirate decay of the free-electron number density $N(t)$ at a fixed temperature T:

$$N(t)=\sum_i N_{i0}\exp(-t/\tau_i) , \tag{5}$$

TABLE II. This table presents the carrier densities and mobilites of van der Pauw samples C and D before and after light exposure at 77 K.

Sample	$N_s(10^{11}$ cm$^{-2})$	Mobility [cm^2/(V s)]
C	3.9 (dark)	65 000 (dark)
	4.97 (after light)	76 000 (after light)
D	3.67 (dark)	39 000 (dark)
	4.47 (after light)	46 000 (atfer light)

where N_{i0} is the initial number density of decay channel i, and with a corresponding individual lifetime τ_i.

Define a quantity $K(t)$, as the ln(t) derivative of $N(t)$, by

$$\begin{aligned} K(t) &= -dN/d[\ln(t)] \\ &= -t\,dN/dt \\ &= \sum_i (t/\tau_i)N_{i0}\exp(-t/\tau_i) . \end{aligned} \tag{6}$$

Notice in Fig. 4 that one decay channel produces a single peak in $K(t)$ (dot-dashed curve) with a width extending over a decade of time. A set of nonvanishing decay features extending over many decades in time indicates a band of decay channels whose characteristic energies and prefactors may be determined by tracking them over a range of T.

Consider the decay process over a narrow enough temperature range T to T' so only one decay mechanism dominates, and one can obtain the following relation between τ_i' and τ_i:

$$\tau_i'=\tau_i\exp[(E/kT')-(E/kT)] , \tag{7}$$

where E is the capture energy, and τ_i' is the new lifetime at T'. Assuming

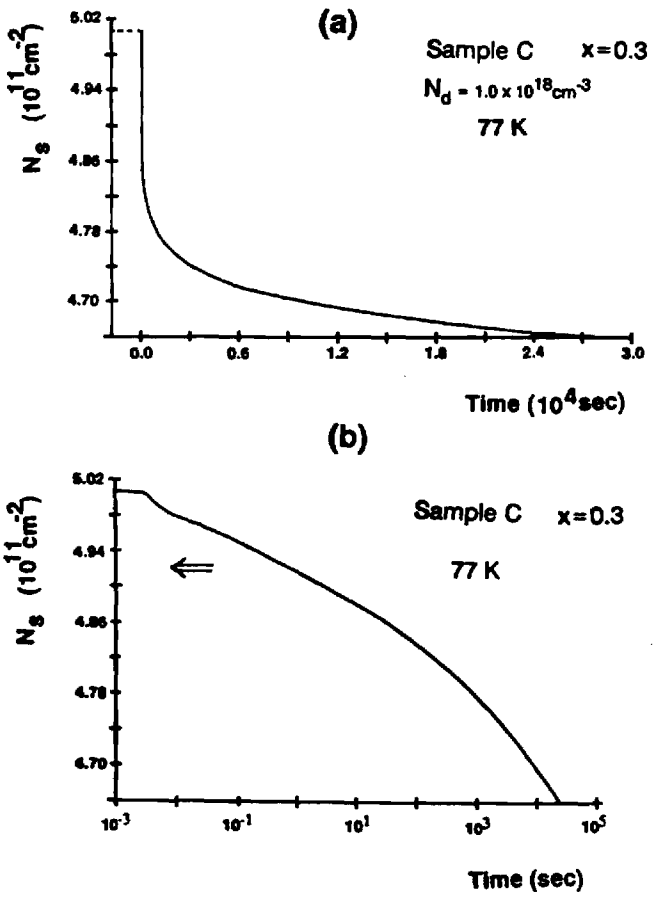


FIG. 3. Multiple-rate TPC decay in the carrier density, N_s from sample C at 77 K. (a) N_s plotted against t. (b) N_s plotted against log(t).

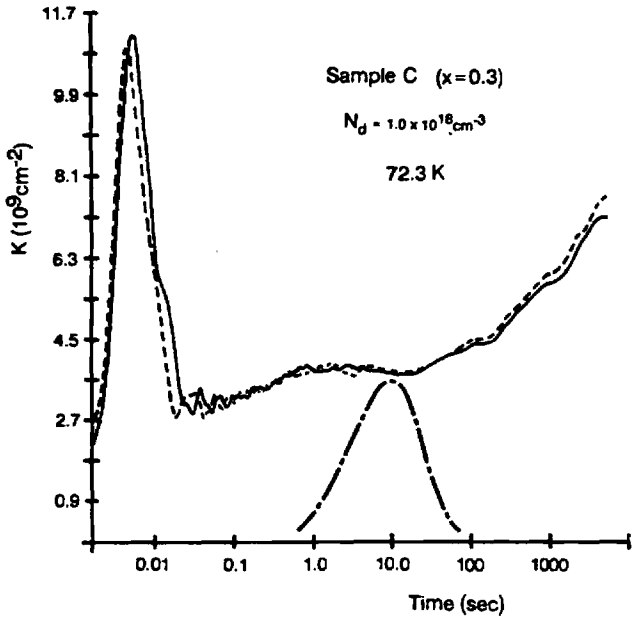


FIG. 4. $K(t)$ of sample C for repeated measurements at 72 K. For comparison, the dot-dashed curve is $K(t)$ for an arbitrary *single* rate decay with $\tau=10$ sec.

$$N'(t)=\sum_i N_{i0}\exp(-t/\tau_i') \tag{8}$$

is the form of the decay at T', the corresponding new value of $K(t)$ at T' is

$$K'(t)=\sum_i (t/\tau_i')N_{i0}\exp(-t/\tau_i') \ . \tag{9}$$

A plot of $K'(t)$ versus $\ln(t)$ will simply be displaced by $[(E/kT')-(E/kT)]$ from $K(t)$ at the same $\ln(t)$. Therefore, the capture energy E_c and the decay time prefactor can be determined from the temperature-dependent displacements of $K(t)$.

When the lifetimes for direct capture into one type of ionized donor and tunneling-assisted capture into another kind of donor are comparable within the same time window, the resulting mixture depends on the amplitude of each decay channel. The fewer the free carriers in the doped $Al_xGa_{1-x}As$ layer after type-A tunneling, the more difficult it is to distinguish a direct capture peak in $K(t)$ from tunneling-assited capture. However the temperature-induced shifts of different peaks or structures will vary if they have different thermal activations.

Although direct capture by DX centers in not very thick doped $Al_xGa_{1-x}As$ layers is difficult to detect, type-B tunneling-assisted capture by DX centers can easily be observed at high temperatures because this directly reduces the 2D electrons. So from Eq. (4), the DX-center capture energy can be deduced indirectly.

V. DETECTION OF PHONON-ASSISTED TUNNELING

This section presents the photodecay data from different samples and identifies the TPC mechanisms discussed in Sec. I. The TPC mechanisms are determined by analyzing the capture energies and lifetime prefactors over different temperature ranges.[14] Raising T speeds up each decay, but for a small enough temperature change, individual peaks or structure can be tracked. The characteristic activation energies of particular decay structures are obtained from their temperature-induced shifts in time.

$K(t)$ was determined from the decay of N_s for van der Pauw samples C and D and the decay of I_{SD} for HEMT sample B. In repeated measurements following illumination (sufficient to saturate I_{SD} or N_s), the $K(t)$ versus $\log_{10}(t)$ plot of the decay curve shows no significant shift of principal features at the same temperature. Data in Fig. 4 shows decay curves, repeated one after another for sample C with identical illumination before each decay. The main features of the data are also reproducible with different histories, such as random thermal cycling or previous illumination at different temperatures.

A. Sample C

Figure 5 shows the temperature dependence of $K(t)$ from sample C for 68 K $< T <$ 85 K. The sample was exposed to 1 sec of white light before each decay. The broad individual decay peaks can easily be tracked on top

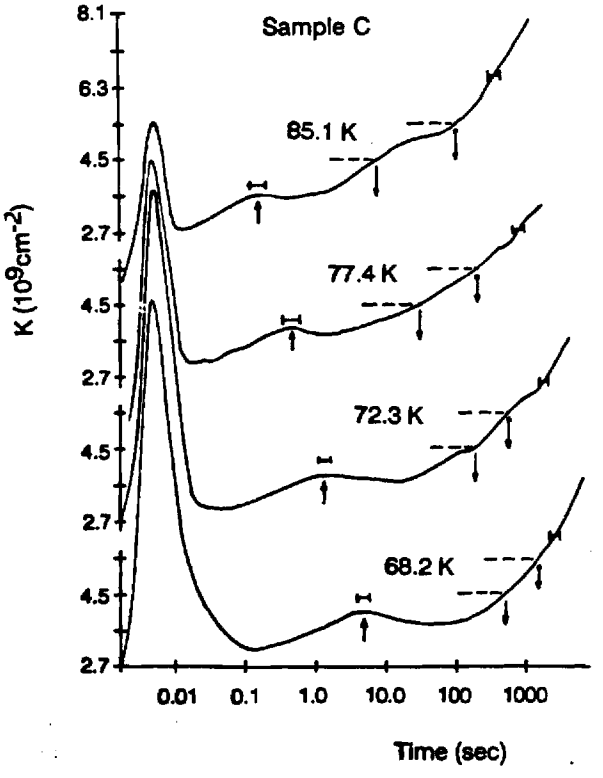


FIG. 5. $K(t)$ for sample C between 68 K $< T <$ 85 K. Up arrows indicate the positions of peaks near $t=1$ sec. Down arrows indicate the positions for specific values of $K(t)$ at each temperature. The horizontal error bars in this and the following figures indicate variations of the positions of either peaks or chosen values of $K(t)$ obtained from different TPC-decay measurements.

of a slowly increasing base line. The error bars indicate the variation for TPC decays repeated at the same temperature. Notice that the peak values of $K(t)$ (near 1 sec) are nearly the same when they shift with temperature. E_c' and τ_0 are deduced from the temperature dependence of the peak shifts. The peaks near 0.01 sec show nearly zero activation energy and $\tau_0 \sim 10^{-2}$ sec. This is identified as direct capture by shallow donors. But the capture energy obtained from peaks near 1 sec is 100 meV with $\tau_0 \sim 10^{-7}$ sec (which is much longer than the DX-center direct-capture prefactor of 10^{-11} sec). For longer times, there are no obvious peaks to track, but a slowly increasing structure is clear, corresponding to a band of decay channels whose overall horizontal shift for the *same* values of $K(t)$ can be tracked at different temperatures. The capture energy of tracks shown in Fig. 5 are measured to be 100 meV±20 meV with $\tau_0 \sim 10^{-3}$–10^{-7} sec. The resulting high thermal activation energy (100 meV) and long-lifetime prefactors are consistent with the prediction in Fig. 2 for type-B tunneling-assisted capture (with a DX capture energy of 80 meV for $T=77$ K).

For higher temperature TPC decay in sample C (86 K $< T <$ 178 K, Fig. 6), the time window is set from 1 msec to 500 sec. In this temperature range, every curve shows a single decay peak (with a few msecs lifetime) that is followed by a slowly increasing baseline. The temperature dependence of the single fast-decay lifetime displays the bahavior of shallow donors seen at lower T (Fig. 5). For 160 $K < T <$ 178 K, a group of peaks are tracked that result in $E_c' = 340 \pm 40$ meV and $\tau_0 \sim 10^{-9}$ sec. This is interpreted as type-B tunneling-assisted capture, with a DX-center capture energy $E_c = 170 \pm 40$ meV, that is consistent with results from other work.[15,16]

B. Sample D

Figure 7 shows the shifts of decay groups with temperature on sample D. The broad individual decay peaks can be easily tracked on top of a slowly increasing base line for 130 K $< T <$ 145 K. Notice also that the peak values of $K(t)$ are nearly the same when they shift with temperature. The peak value of $K(t)$ is $\approx 1 \times 10^{10}$ cm^{-2}, which corresponds to a decrease in N_s of 1.2×10^{10} cm^{-2}. E_c' is measured to be 400±40 meV with $\tau_0 \sim 10^{-13}$ sec. This high apparent capture energy for 130 K $< T <$ 145 K is interpreted as type-B tunneling-assisted capture, and the resulting value of E_c is 230±40 meV, which is slightly higher than the results from samples C and L with the same Al fraction.

By contrast, in Fig. 8 (sample D) for 82 K $< T <$ 110 K, there are no obvious peak shifts to track, but a slowly increasing base-line structure becomes clear, whose overall horizontal shift at the same values of $K(t)$ can be tracked as a function of temperature. In addition note for comparision the superposition (dashed line) of the resulting $K(t)$ peak for a theoretical single rate decay process with $\tau_0 \sim 100$ sec. The capture energy of the experimental tracks is measured to be about 20 meV with $\tau_0 \sim 1$–100 sec. This lower activation energy and long-lifetime prefactor together with an increasing base-line structure is not explained by the results shown in Fig. 2.

The behavior of $K(t)$ for sample D was verified in an alternative decay experiment. A continuous temperature sweep at a constant rate of about 0.1 K/sec after il-

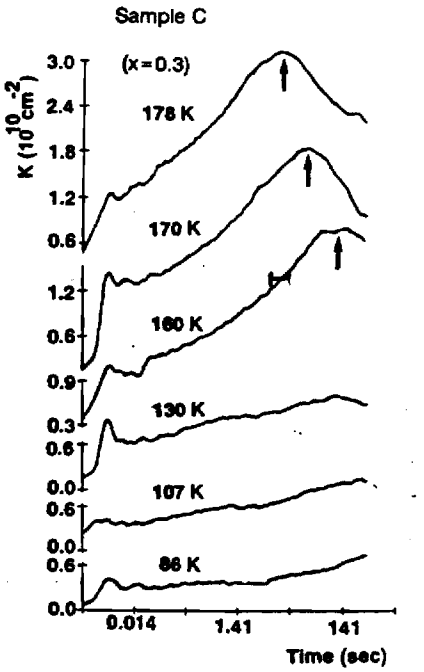


FIG. 6. $K(t)$ from sample C for 86 K $< T <$ 178 K. A 5 sec LED light pulse is used for illumination, with the data showing no significant change from exposure to white light.

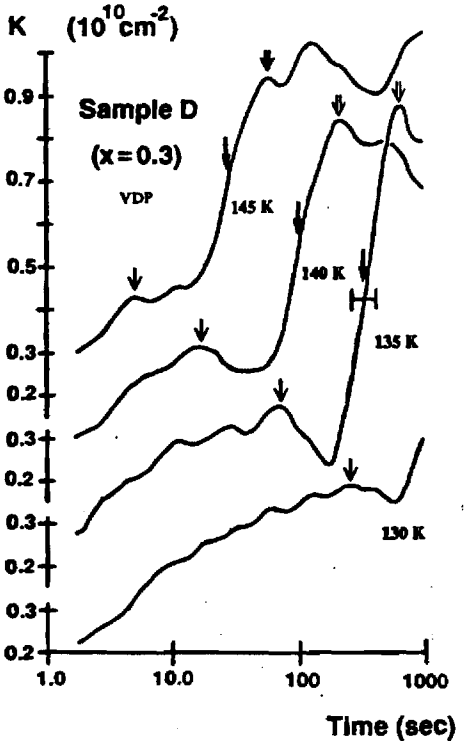


FIG. 7. $K(t)$ of sample D for 130 K $< T <$ 145 K. Tracked peaks are indicated by short arrows and the "shoulders" are indicated by long arrows.

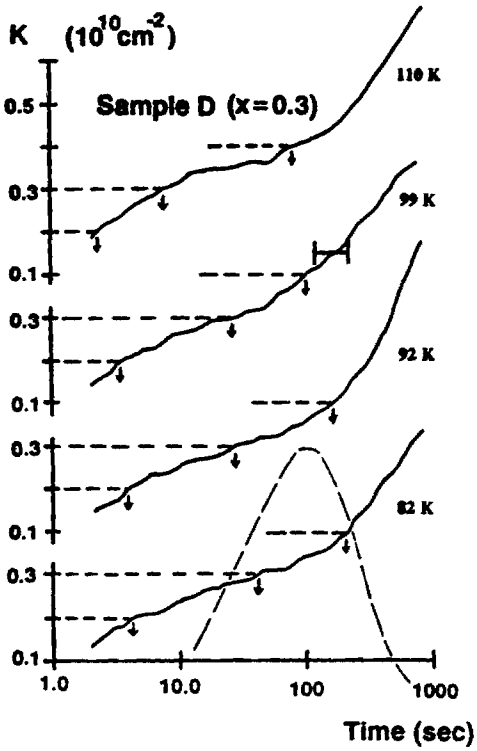


FIG. 8. $K(t)$ of sample D for 82 K $< T <$ 110 K. In comparison, the dashed curve is $K(t)$ for a single rate decay with $\tau = 100$ sec. Down arrows indicate the time positions for chosen values of $K(t)$.

lumination is shown in Fig. 9. When the sample is warmed from 77 K, a sudden dip in carrier density is observed (see arrow) when T passes through 150 K, which agrees with the temperatures where large-$K(t)$ peaks were observed in Fig. 7. The decrease in N_s ($\approx 1.5\times10^{10}$ cm^{-2}) is in good agreement with the value (1.2×10^{10} cm^{-2}) obtained from the peak value of $K(t)$ in Fig. 7.

C. Sample B

Figure 10 is an example of TPC decay in sample B (from conductivity measurements) from 55 K $< T <$ 74 K. $K(t)$ displays a monotonically increasing base-line structure and has almost zero thermal activation energy. The monotonic behavior of $K(t)$ indicates that the decay is occurring from a range of decay channels over a continuous spread of relaxation rates. At zero gate voltage (V_g), there are some free electrons from shallow donors ($\sim$20% of the total donors for the case of $x = 0.3$) in the thick (100 nm, much wider than W_d in-dark) doped $Al_xGa_{1-x}As$ layer. E_{CB} will line up with E_0 of the 2D EG (unlike the case shown in Fig. 1), so that no type-B tunneling-assisted decay takes place. After type-A tunneling, thermally activated behavior by direct capture to deep centers over a narrow intermediate temperature range 60 K $< T <$ 80 K was expected, considering the results of other work on bulk material[17] and by decay of the threshold voltage which is sensitive to direct capture to deep centers in the $Al_xGa_{1-x}As$ layer.[15] However the results of this work show that the carrier density in the $Al_xGa_{1-x}As$ layer was not sufficient to produce a signal from direct capture that was large enough to be clearly detected from the decay background.

Figure 11 shows $K(t)$ from sample B for 122 K $< T <$ 159 K ($V_g = 0$) resulting in $E_c = 170\pm20$ meV and $\tau_0 \sim 10^{-6}$ sec. This energy is too low to agree with the apparent capture energy ($E_c' = 340$ meV) of type-B tunneling-assisted decay in the same temperature range. This is consistent with the fact that type-B tunneling-assisted capture is not expected in this case where in a

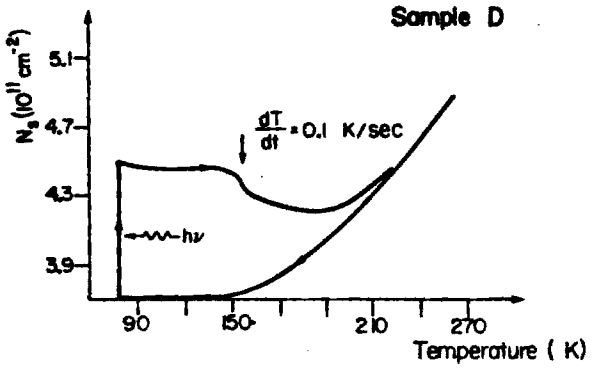


FIG. 9. Temperature-sweep method of observing elevated TPC decay at special temperatures (arrow) for sample D.

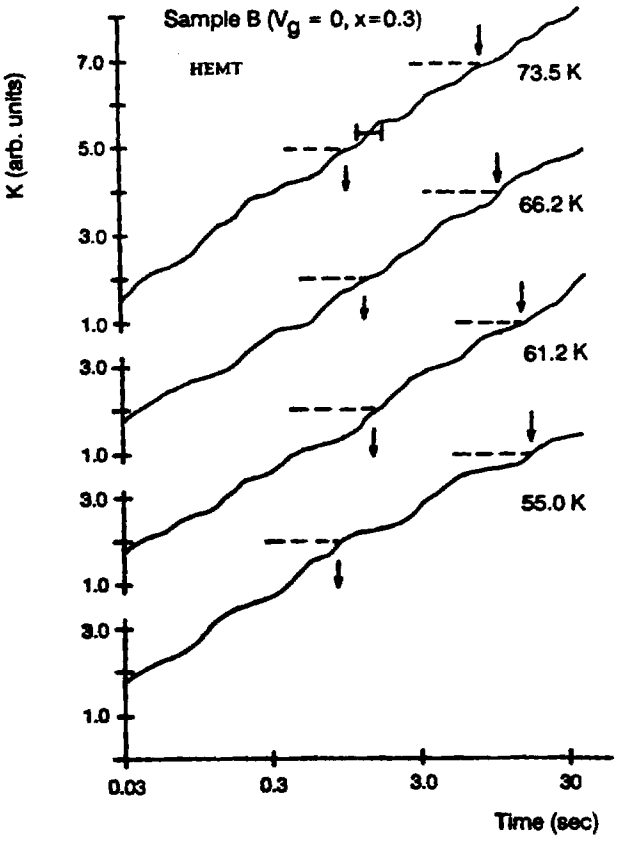


FIG. 10. $K(t)$ of HEMT sample B for 55 K $< T <$ 74 K ($V_g = 0$). Down arrows indicate the time positions for chosen values of $K(t)$.

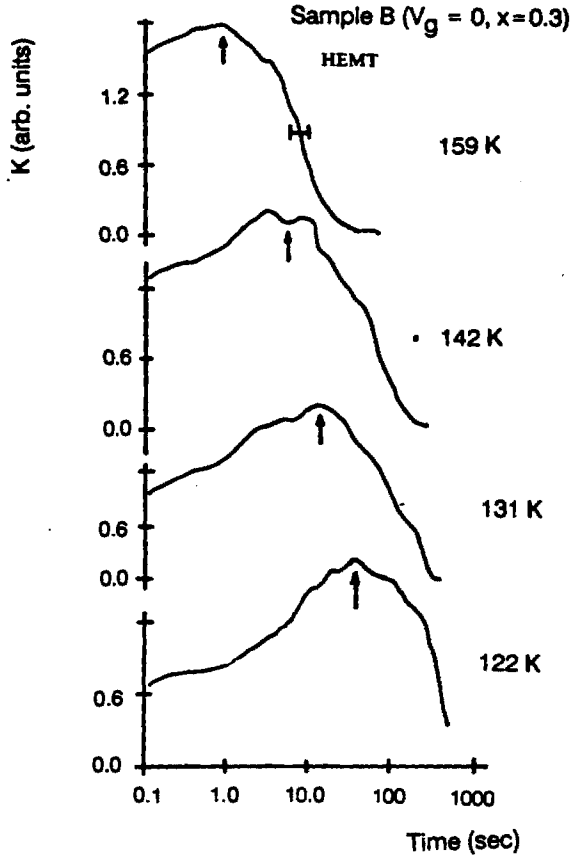


FIG. 11. $K(t)$ of HEMT sample B for 122 K $< T <$ 159 K ($V_g = 0$).

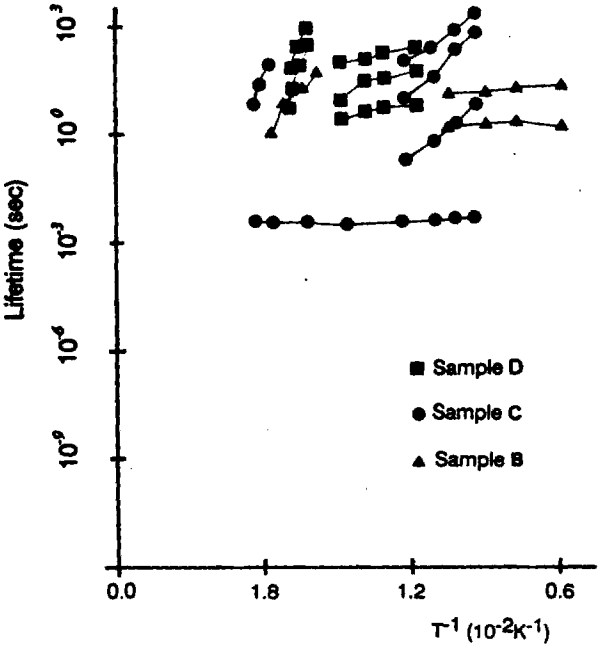


FIG. 12. Summary of the tracked peaks or other $K(t)$ structures. The lines for sample D (82 K $< T <$ 110 K) and HEMT sample B (55 K $< T <$ 74 K), and the upper two lines from sample C (68 K $< T <$ 85 K) are obtained from the horizontal shifts of chosen values of $K(t)$. Data connected by solid lines are consistent with the picture represented in Fig. 2, whereas data connected by dashed lines are not.

thick doped $Al_xGa_{1-x}As$ layer (100 nm) E_{CB} does not rise above E_0.

VI. SUMMARY

Using an analysis method based on $dN/d[\ln(t)]$ (or alternatively $dI_{SD}/d[\ln(t)]$) we have been above to detect distinct decay features of transient photoconductivity in $GaAs/Al_xGa_{1-x}As$ heterostructures. This method was employed on different heterostructures and HEMT devices. The results permitted us to identify the contribution to TPC of phonon-assisted tunneling from other PPC-decay mechanisms.

The thermal activation features described in the previous section are summarized in Fig. 12. The results, which are consistent with the predictions in Fig. 2, are connected by solid lines. The data shows TPC decay occuring mainly through direct capture by shallow traps and type-B tunneling-assisted capture by deep traps in the $Al_xGa_{1-x}As$ layer. At temperatures near 77 K, only deep-trap-related type-B tunneling-assisted decay (with $E_c = 80 \pm 20$ meV) was observed within the time window 10^{-3}–10^{4} sec. At temperatures near 160 K, trap-related type-B tunneling-assisted decay dominates (with capture energy between 170–230 meV).

The DX-center capture energies ($E_c = 80$–230 meV) for $x = 0.3$ deduced from type-B tunneling-assisted decay are in reasonable agreement with results ($E_c = 110$–200 meV) from other work.[14,15] Direct capture by deep traps in the $Al_xGa_{1-x}As$ layer was not observed due to the low mobility and small number of free electrons in this layer, hence their relatively small contribution to the transport current compared to that of the 2D EG.

Other data connected by dashed lines are not consistent with the picture in Fig. 2. For example, low activation energy ($E_c < 20$ meV) features are present for 50 K $< T <$ 110 K but with lifetime prefactors longer ($\tau_0 > 1$ sec) than the tunneling calculation of Fig. 2 would permit. Also, thermally activated behavior with $E_c = 170$ meV and $\tau_0 \sim 10^{-6}$ sec is present for 120 K $< T <$ 160 K in HEMT sample B ($V_g = 0$). This capture energy is too small to agree with type-B tunneling-assisted decay at the same temperatures. All of the above long prefactors indicate the involvement of tunneling-assisted decay. These features are consistent with the experimental results[3] which led to the proposal of 2D-electron virtual-state impurity tunneling. Since the proposed virtual states are not yet clarified, more theoretical modeling and experimental work are needed to determine its contribution to TPC.

ACKNOWLEDGMENTS

This work was supported by the NSF-DMR Grant Nos. 81-19550, 81-13456, and 85-19728 and by the Tektronix Foundation. It is a pleasure to thank K. R. Gleason (Tektronix) and W. Schaff (Cornell) for supply-

ing useful samples and J. D. Cohen and J. C. Overley (University of Oregon) for helpful discussions.

APPENDIX

Expressions for the tunneling barrier $U(z)$, which begins from the level E_{th} above E_0, and the tunneling barrier width z_0 are calculated as follows for conduction-band bending [shown as the solid cruve in Fig. 1(a)] after illumination. $U(z)$ and z_0 are determined from E_{th}, ΔE_c, and L as shown in Fig. 1(a), and the ionized donor number density N_i (in units of 10^{18} cm^{-3}). All energies are in units of eV, and all distances are in nm.

The value of the depletion width after type-A tunneling ($W_d = z_1 - L$) is obtained by

$$\Delta E_c - E_0 = [(z_1 - L)^2 + 2L(z_1 - L)]N_i/1458 . \quad \text{(A1)}$$

Given a value of the depletion width ($W_d = z_2 - L$) at this moment with band bending shown as the solid curve in Fig. 1(a), the difference between the $Al_xGa_{1-x}As$ layer conduction-band minima and E_0 of 2D EG is

$$\Delta E = [(z_2 - L)^2 + 2L(z_2 - L)]N_i/1458 - \Delta E_c + E_0 , \quad \text{(A2)}$$

and the value of U_0 shown in Fig. 1(a) is

$$U_0 = (z_2 - L)^2 N_i/1458 . \quad \text{(A3)}$$

Then, in the case of $U_0 - \Delta E > E_{th}$, i.e., $z_0 > L$, we have

$$z_0 = z_2 - [(\Delta E + E_{th})1458/N_i]^{1/2} . \quad \text{(A4)}$$

There are two formulas for $U(z)$. If $z < L$,

$$U(z) = (\Delta E_c - E_0 - U_0 + \Delta E)(1 - z/L) + U_0 - \Delta E - E_{th} ; \quad \text{(A5)}$$

but if $z > L$,

$$U(z) = U_0[1 - (z - L)/(z_2 - L)]^{1/2} - \Delta E - E_{th} . \quad \text{(A6)}$$

In the case of $U_0 - \Delta E < E_{th}$, i.e., $z_0 < L$, there are different formulas:

$$z_0 = L(\Delta E_c - E_0 - E_{th})(\Delta E_c - E_0 - U_0 + \Delta E)^{-1} \quad \text{(A7)}$$

and

$$U(z) = (\Delta E_c - E_0 - U_0 + \Delta E)(1 - z/L) + U_0 - \Delta E - E_{th} . \quad \text{(A8)}$$

The above equations can be applied to type-B tunneling by replacing z_2 and U_0 by their counterparts z_2^* and U_0^* [Fig. 1(b)].

---

*Corresponding author.

[1]For a recent review see M. I. Nathan, Solid-State Electron. **29**, 167 (1986).

[2]D. E. Theodorou and H. J. Queisser, J. Appl. Phys. **23**, 121 (1980).

[3]E. F. Schubert, A. Fischer, and K. Plogg, Phys. Rev. B **31**, 7937 (1985).

[4]D. V. Lang and R. A. Logan, Phys. Rev. Lett. **39**, 635 (1977).

[5]D. V. Lang and R. A. Logan, Phys. Rev. B **19**, 1015 (1979).

[6]K. L. I. Kobayashi, Y. Uchida, and H. Nakashima, Jpn. J. Appl. Phys. **24**, L928 (1985).

[7]H. J. Queisser and D. E. Theodorou, Phys. Rev. **33**, 4027 (1986).

[8]E. F. Schubert and K. Ploog, Phys. Rev. **29**, 4562 (1984).

[9]E. F. Schubert, J. Knecht, and K. Ploog, J. Phys. C **18**, L215 (1985).

[10]N. Chand, T. Henderson, J. Klem, W. T. Masselink, R. Fischer, and H. Morkoç, Phys. Rev. B **30**, 4481 (1984).

[11]The pure-tunneling lifetime can be estimated as $Wv^{-1}p^{-1}$, where W is the width of the potential well at the 2D interface, v is the z-direction velocity of the 2D electrons deduced from the z-quantized ground-state energy E_0 plus a thermal energy in the case of phonon-assisted tunneling. Using as estimate of $W \sim 10^{-6}$ cm, and $v \sim 10^8$ cm sec^{-1} along with the range of pure-tunneling probabilities (10^{-6}–10^{-10}) from the WKB integral, one gets tunneling lifetimes between 10^{-8} to 10^{-4} sec.

[12]O. Madelung, *Introduction to Solid-State Theory* (Springer, New York, 1978), p. 177.

[13]L. X. He, K. P. Martin, and R. J. Higgins, Phys. Rev. B **36**, 6508 (1987).

[14]For convenience, the value of $K(t)$ obtained from this data is with respect to $\log_{10}(t)$, which is a factor of 2.30 different than the value defined by Eq. (6). In this way, a single-rate decay with an amplitude of N_∞ corresponds to an experimental result of $K(t)$ with a peak value of $0.85N_\infty$.

[15]P. M. Mooney, Bull. Am. Phys. Soc. **32**, 504 (1987).

[16]P. M. Mooney, N. S. Caswell, P. M. Solomon, and S. L. Wright, Mater. Res. Soc. Symp. Proc. **46**, 403 (1985).

[17]R. J. Nelson, Appl. Phys. Lett. **31**, 351 (1977).

Infrared quenching of persistent photoconductivity in GaAs/$Al_xGa_{1-x}As$ heterostructures

L. X. He
Department of Physics, University of Oregon, Eugene, Oregon 97403

K. P. Martin and R. J. Higgins
School of Electrical Engineering and Microelectronics Research Center, Georgia Institute of Technology, Atlanta, Georgia 30332

(Received 5 January 1988; revised manuscript received 11 July 1988)

Carrier number density and conductance measurements on GaAs/$Al_xGa_{1-x}As$ heterostructures for $2<T<77$ K confirm there is an infrared-induced (ir, $h\nu<0.7$ eV) quenching of persistent photoconductivity (PPC) in samples with $x=0.30$, 0.23, and 0.21. This is observed in both ungated and gated heterostructures with small gate biases. Measurements of de Haas–Shubnikov (dHS) oscillations show a decrease (relative to the PPC state) of the two-dimensional electron number density from the ir illumination, accompanied by increases in the dHS oscillation amplitude and the carrier scattering lifetime when the $Al_xGa_{1-x}As$ layer is not fully depleted. Together with shifts of the transconductance peak gate voltage and threshold voltage during ir quenching, the data establish that the decay in PPC is dominated by the capture of ir-excited free electrons back to deep centers in the $Al_xGa_{1-x}As$ layer. The capture-cross-section value estimated from the data is about 10^{-20} cm^2, which agrees with a calculation based on the energy of the ir-excited free electrons and the deep-trap-capture barrier height and decay prefactor. Data from gated samples show a gate electric field effect, where ir quenching (low electric fields) changes to ir enhancement (strong electric fields). This is discussed in terms of an electric-field-induced decrease in the optical threshold of emission from *DX* centers.

I. INTRODUCTION

Recently, we and others[1] have observed that while broadband white light ($h\nu>0.8$ eV) enhances conductivity in high-electron-mobility transistor (HEMT) structures, infrared light (ir, $h\nu>0.5$–0.8 eV), can erase a portion of the persistent-photocurrent effect (PPC) enhanced carrier density. An understanding of this phenomenon, known as infrared quenching (IRQ), helps resolve remaining puzzles of persistent photoconductivity and offers an opportunity for an intriguing application of PPC for optical information storage.

Nathan *et al.*[1] have observed infrared ($h\nu=0.6$–0.8 eV) quenching (as a 2% increase in the resistance) in an ungated GaAs/$Al_xGa_{1-x}As$ heterostructure sample ($x=0.3$) at 80 K. It was suggested[1] that this increased resistance occurs when the electrons are excited out of the high-mobility two-dimensional electron gas (2D EG) by absorption of the ir light and cross the barrier into the $Al_xGa_{1-x}As$ where they are recaptured. While the macroscopic barrier present in GaAs/$Al_xGa_{1-x}As$ heterostructures ordinarily prevents PPC-induced electrons at the 2D interface from rapid recapture by ionized donors in the $Al_xGa_{1-x}As$ layer, an ir-excited free electron can easily overcome this barrier (≈ 0.2 eV). The details were not explored in that work, but it may be noted that the photon energies where IRQ was observed (0.6–0.8 eV) were not sufficient to make electrons hop macroscopic recapture barriers (>0.8 eV) to recombine with holes in the epitaxially grown GaAs buffer layer or ionized impurities in the GaAs substrate.

Obviously free-carrier absorption of ir photons can occur in the Si-doped $Al_xGa_{1-x}As$ layer. But no reports have appeared that show significant infrared quenching of PPC in essentially "bulk" $Al_xGa_{1-x}As$ material which is generally grown on an undoped GaAs substrate and may actually be regarded as a GaAs/$Al_xGa_{1-x}As$ heterojunction with a thick (~ 1 μm) doped $Al_xGa_{1-x}As$ layer.[2] This weak ir response in "bulk" doped $Al_xGa_{1-x}As$ is consistent with the *DX*-center model shown in Fig. 1. The occupied *DX* center (trapped-carrier state) has an energy above the conduction band when at the configuration coordinate value where the unoccupied *DX* center (free-carrier state) sits with minimum energy. A valence electron, whose energy level is at least E_{gap} below the occupied *DX* center, cannot be optically transferred to a *DX* center by infrared light ($h\nu<E_{gap}$). However, as suggested by Nathan *et al.*, a conduction-band electron (i.e., unoccupied *DX*-center configuration) can be recaptured, forming an occupied *DX* center when the free electrons are excited by infrared light with photon energy E_1 smaller than the onset photon energy $E_{optical}$ (Fig. 1).

This *DX*-center filling mechanism is expected to be a relatively weak effect in bulk $Al_xGa_{1-x}As$ due to a conduction electron's small optical-absorption cross section (shown to be 10^{-19} cm^2 in the Appendix). The change in the $Al_xGa_{1-x}As$ carrier density (N_1) in HEMT structures is also expected to be small in measurements of the total sheet number density N_s or source-drain current, I_{SD} [Eqs. (1) and (2)]. The charge control model[3] and parallel conductance experiments[4] have shown that N_s and I_{SD} are dominated by the 2D EG density (N_{2D}) whose mobility is 100 times larger than carriers in the

$Al_xGa_{1-x}As$. However, a multiplier effect comes into play when the $Al_xGa_{1-x}As$ layer is depleted, and N_{2D} becomes sensitive to small changes in the ionized donor density of the doped $Al_xGa_{1-x}As$ layer (N_i). For example, a 1% decrease in a nominal 2×10^{17} cm^{-3} ionized donor density in a 100-nm-thick doped $Al_xGa_{1-x}As$ layer changes N_{2D} by 2×10^{10} cm^{-2} out of a total 4×10^{11} cm^{-2} (e.g., Fig. 2) or 5%, for a multiplier effect of 5. For HEMT samples with the gate voltage (V_g) near threshold (V_t), the multiplier effect can be much bigger because of the much smaller total N_{2D}. We propose that this is why IRQ is observed to be most significant for samples without free carriers in the $Al_xGa_{1-x}As$ layer.

The following sections explore several aspects of the capture of ir-excited free electrons. (i) When infrared light with $h\nu<0.65$ eV is used (thus avoiding contributions from the GaAs buffer layer and substrate), it is proposed that IRQ occurs by the capture of ir-excited electrons into traps in the doped $Al_xGa_{1-x}As$. Within this capture picture N_i decreases, with a corresponding decrease in N_{2D}. These changes in N_i and N_{2D} also change the 2D EG mobilities, which can be measured by either classical or quantum (de Haas–Shubnikov) methods.[5] (ii) The excitation is due to absorption of ir photons by free electrons. The capture cross section of ir-thermalized electrons trapped by deep levels must agree with the deep-trap-capture barrier heights and decay prefactors determined from independent thermal capture experiments. The discussion of part (ii) is left for Sec. VI. Experimental results from a variety of gated and ungated samples from different sources will be shown to be consistent with the above picture.

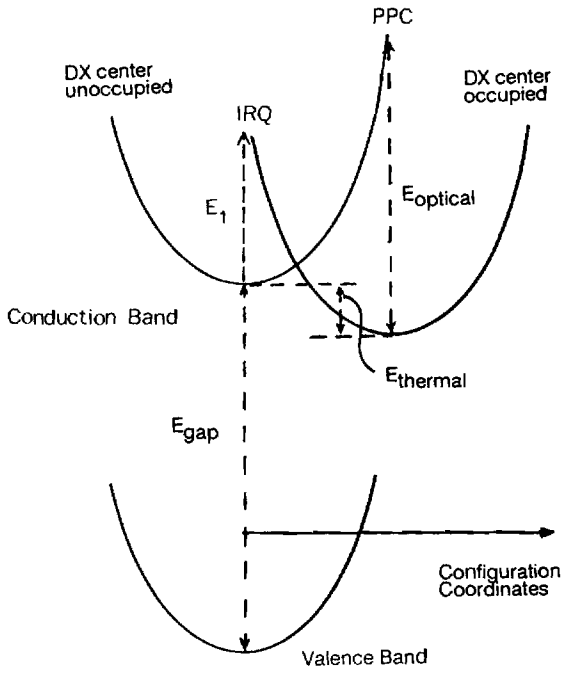


FIG. 1. The configuration coordinate model for a *DX* center. The figure shows the path for PPC onset and transition energy $E_{optical}$, as well as that for ir quenching and onset threshold, E_1.

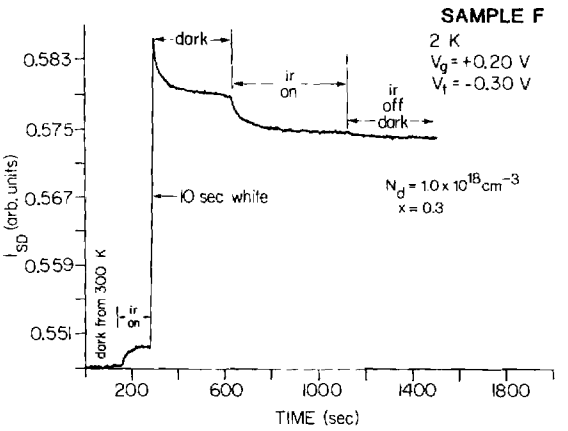


FIG. 2. Infrared-quenching: A typical source-drain current from sample F at 2 K under the influence of brief white-light and continuous ir illumination.

II. SAMPLES AND MEASUREMENTS

Most of the molecular-beam-epitaxy (MBE) grown heterostructures used in this work had the following layer sequence: Si-doped GaAs cap layer, Si-doped $Al_xGa_{1-x}As$ layer, undoped $Al_xGa_{1-x}As$ spacer layer, undoped GaAs buffer layer, and semi-insulating GaAs substrate. A 2D EG forms at the interface between the undoped $Al_xGa_{1-x}As$ and GaAs buffer layers, separated by the spacer layer from the nearest ionized donors in the $Al_xGa_{1-x}As$ layer. A typical device, the so-called high-electron-mobility transistor, has a metal gate on top of the GaAs cap layer with a length (L) to width (W) ratio $L/W>10$. Another type of device is a simple van der Pauw (VDP) square with an Ohmic contact in each corner.

Samples from four different MBE sources were used. Summarized in Table I are their principal parameters: aluminum fraction x, cap-layer thickness L_c, doping density $N_{d,c}$ of the cap layer, $Al_xGa_{1-x}As$ layer thickness L_a, $Al_xGa_{1-x}As$ layer doping density N_d, undoped spacer thickness L_s, gate voltage threshold (V_t) at 77 K in the dark, the net ionized donor density N_i in the depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, and the percentage increase of N_i after identical brief white light illumination sufficient to saturate the induced PPC. Samples A–C, E, and F were gated samples. Samples A, B, and G are on the same chip, but sample G is ungated. Samples A, B, F, and G had the same Al fraction ($x=0.3$) while samples C and E had $x=0.21$. Sample D, an ungated VDP, had $x=0.3$ in the doped $Al_xGa_{1-x}As$ layer but $x=1.0$ in the undoped spacer layer. Sample H is an ungated VDP square with $x=0.23$.

For gated HEMT samples, the source-drain current I_{SD} was measured with a fixed source-drain voltage (V_{SD}). For ungated Van der Pauw samples, N_s was measured via the Hall effect. When the 2D EG mobility (μ_{2D}) is much larger than the mobility of electrons in $Al_xGa_{1-x}As$ layer (μ_1) such that $\mu_1N_1 \ll \mu_{2D}N_{2D}$ (the difference between Hall mobility and conductance mobility is ignored because only the ratio μ_1/μ_{2D} is important

TABLE I. Principal sample parameters. Sample label, source, aluminum ratio x, cap-layer thickness L_c, doping density $N_{d,c}$ of the cap layer, $Al_xGa_{1-x}As$ layer thickness L_a, $Al_xGa_{1-x}As$ layer doping density N_d, undoped spacer thickness L_s, gate voltage threshold (V_t) at 77 K in the dark, the net ionized donor density N_i in the depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, and the percentage increase of N_i after identical brief white-light illumination sufficient to saturate the PPC-induced change. The layer thicknesses are in units of nm and doping or carrier densities are in units of 10^{17} cm^{-3}.

Sample	Source	x	L_c (nm)	$N_{d,c}$	L_a (nm)	N_d	L_s (nm)	V_t (V)	N_i	$\Delta N_i/N_i$
A,B	Tek[a]	0.3	50	2	100	10	5	−0.5	1.9	15%
C	TCSF[b]	0.21	20	6	120	6	8	−2.0	2.9	11%
D	Cornell[c]	0.3	20	10	30	10	10		ungated VDP	
G	Tek[a]	0.3	50	2	100	10	5		ungated	
E	TCSF[b]	0.21	20	6	120	6	8	−1.2	2.3	13%
F	Tek[a]	0.3	50	2	100	10	5	−0.2	1.6	20%
H	GE[d]	0.23	25	10	10	10	15		ungated VDP	

[a]Tek: Tektronix, Inc.
[b]TCSF: Thomson-CSF (Paris).
[c]Cornell: Cornell University.
[d]GE: General Electric Company.

in this case), N_s and I_{SD} can be written as

$$N_s = N_{2D} + 2N_1\mu_1/\mu_{2D} \quad (1)$$

(Hall measurement[6]) and

$$I_{SD} = -eGV_{SD}\mu_{2D}(N_{2D} + N_1\mu_1/\mu_{2D}) \quad (2)$$

(conductivity measurement) where G $(=W/L)$ is a geometrical factor for short-gate samples. In this study, $\mu_1 \approx 10^3$ cm^2/V sec while $\mu_{2D} \approx 10^5$ cm^2/V sec and N_1 never exceeds N_{2D} significantly, such that the condition for Eq. (1), $\mu_1 N_1 \ll \mu_{2D} N_{2D}$, is satisfied.

The ir light came from a white-light source filtered through a germanium lens with a cutoff at 0.67 eV. This is well below the onset optical threshold of 0.8 eV for PPC in $GaAs/Al_xGa_{1-x}As$ heterostructures.[1] Infrared quenching of PPC was observed for $2 < T < 77$ K in all samples. Figure 2 shows a sequence of I_{SD} versus time and light exposure where sample F has first been cooled from 300 to 2 K in the dark. The initial response to ir light is a weakly enhanced conductivity, in contrast to the larger increase in I_{SD} from white-light illumination. When the white light was removed, I_{SD} approaches a base line far above the initial in-dark value following an initial fast decay. After this, exposure to infrared illumination partially quenches the PPC enhancement of I_{SD}. Under steady ir illumination the quenching begins

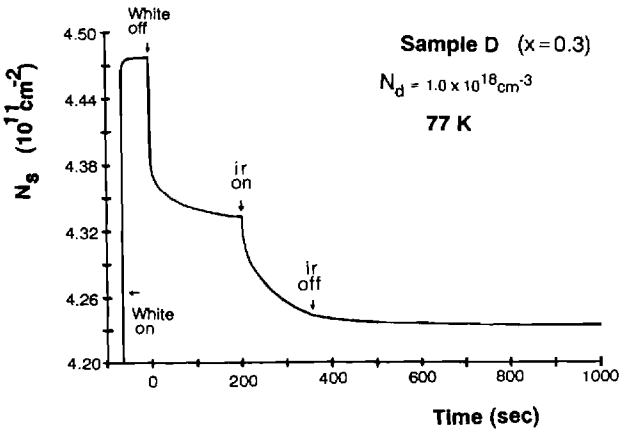


FIG. 3. Measurement of N_s during ir quenching in sample D.

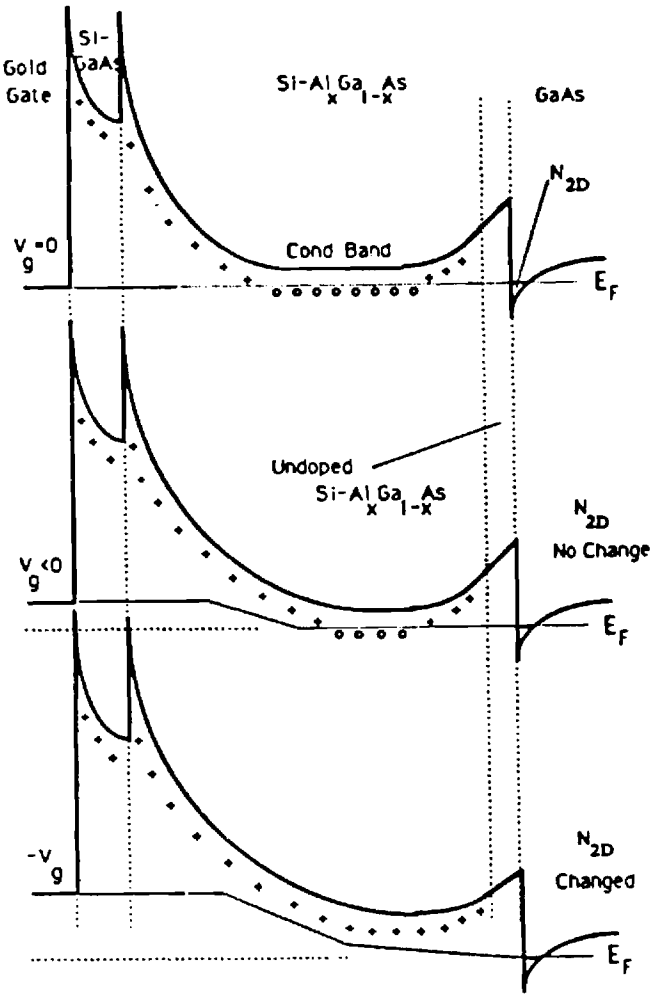


FIG. 4. The conduction-band bending in different layers of a HEMT device under different gate biases, going from (top) parallel conductance in the $Al_xGa_{1-x}As$ layer ($V_g = 0$) to (bottom) full depletion in the $Al_xGa_{1-x}As$.

with a fast relaxation that slows as it approaches a base line value. For later times at these temperatures I_{SD} remains virtually unchanged after ir illumination is removed.

The base line value from IRQ does *not* depend on the illumination intensity. But the quenching decay time is strongly dependent on the ir light intensity. For example, when the ir intensity was increased by a factor of 10, the quenching time to the base line value was also reduced by a factor of 10.

At 77 K IRQ decreases N_s by 2% in sample D (Fig. 3) and 5% in sample H. For HEMT samples with V_g set so there is parallel conductance in the $Al_xGa_{1-x}As$ layer from low-mobility carriers [Fig. 4, (top)], the relative decrease of I_{SD} from IRQ is $\sim 1\%$. But when V_g is set close to V_t and only the 2D EG remains [Fig. 4, (bottom)], IRQ reduces I_{SD} by more than a factor of 10.

III. INFRARED QUENCHING OF PERSISTENT PHOTOCONDUCTIVITY

Since ir quenching reduces I_{SD} (Fig. 2), it is clear that the mechanism does not transfer electrons from the $Al_xGa_{1-x}As$ layer to the 2D interface, since $\mu_{2D} \gg \mu_1$ so I_{SD} would increase rather than decrease. To determine the net change in carriers in the $Al_xGa_{1-x}As$ layer, information other than simply the change in the source-drain current or overall sheet number density is required.

A. Transconductance of gated samples

For the gated depletion mode devices used, the conduction band bending across the $Al_xGa_{1-x}As$ layer was changed by varying the gate bias (Fig. 4). According to the charge control model for HEMT devices,[3,7] initial decreases (more negative) in V_g reduce the conductance of the $Al_xGa_{1-x}As$ layer while N_{2D} remains unchanged. When all carriers in the $Al_xGa_{1-x}As$ layer are depleted by V_g, N_{2D} begins decreasing with further lowering of V_g until channel pinchoff occurs. From this picture of charge control, the transconductance ($g_m = dI_{SD}/dV_g$) shows low values in the range of V_g which controls just the low-mobility carriers in the $Al_xGa_{1-x}As$ layer, but reaches much higher values when the gate bias controls the high-mobility 2D EG. When the transconductance is measured from $V_t < V_g < 0$, the g_m peak occurs at $V_g = V_p$, just past the point of $N_1 = 0$, and V_t occurs when N_1 and N_{2D} have been depleted. Therefore, shifts of V_p and V_t during thermal or optical processes provide important information about the carrier density changes in different heterostructure layers.

Figure 5 shows the following sequence of transconductance curves for sample A at 77 K: (1) sample initially in the dark, (2) after 30 sec of exposure to white light, (3) 1 min after removing the white light, and (6) after 2 min of infrared-light exposure. In addition, portions of the transconductance peak are shown (4) immediately after ir illumination begins and (5) 30 sec later during IRQ. The data in Fig. 5 show a shift of V_p from -0.15 to -0.42 V when going from the dark state to a fully saturated PPC state. V_p relaxes back to -0.38 V after the light is removed and then shifts further to -0.28 V upon receiving

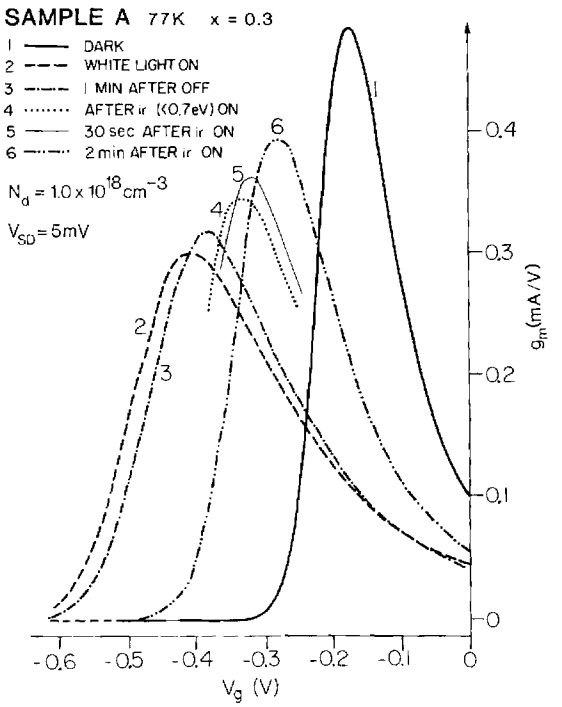


FIG. 5. Typical transconductance peak shifts under the influence of brief white light and continuous ir illumination from sample A at 77 K.

infrared illumination. V_t shows similar behavior. The relative difference between V_t and V_p also changes during these processes (0.13 V in dark, 0.17 V with white light on, and 0.15 V after IRQ), which indicates changes in depletion of the 2D EG. The change in N_1 at $V_g = 0$ is estimated from shifts of V_t or V_p by assuming the capacitance to be $C \approx \epsilon A/d$.[3] For sample A $\epsilon = 13$ is the dielectric constant, $A = 3 \times 10^{-6}$ cm^2 the area of gate, and $d = 150$ nm the distance between the gate and 2D EG interface with the result $C = 0.3$ pF. From $\Delta Q = C\,\Delta V_t$, there is a 3×10^{16}-cm^{-3} increase in N_1 upon white-light illumination, and a 1×10^{16}-cm^{-3} decrease upon ir illumination, for a 100-nm-thick doped $Al_xGa_{1-x}As$ layer. The estimated changes in N_1 (+3% with white light on, -1% after IRQ) are consistent with the relative changes in I_{SD} (+6% with white light on, -1% after IRQ) compared to $N_d = 1 \times 10^{18}$ cm^{-3} for HEMT samples with large parallel conductance in the $Al_xGa_{1-x}As$ layer (Fig. 2).

A factor of 2 decrease in the g_m peak amplitude is observed in the PPC state followed by the recovery of higher g_m values with IRQ (Fig. 5). This is evidence for N_i increasing in the PPC state and decreasing from IRQ. Screening effects are of secondary importance since N_{2D} only changes by a few percent.

Sample E with a lower in-dark threshold voltage of -1.2 V at 77 K shows a similar pattern: V_p shifted from -1.1 V (in dark) to -1.4 V (white light) and later decays to -1.35 V during the relaxation following white-light illumination and shifts further to -1.15 V upon exposure to infrared illumination.

The behavior of the transconductance permits us to conclude unambiguously that the free-carrier density or

the ionized donor density in the $Al_xGa_{1-x}As$ layer increases upon white-light illumination but later decreases upon receiving ir illumination.

B. de Haas–Shubnikov oscillations

Considerable evidence of the proposed ir-quenching mechanism was also accumulated from the de Haas–Shubnikov (dHS) measurements. The dHS oscillation amplitude σ_{OSC} at low magnetic fields measures the 2D EG mobility μ_{2D} ($=e\tau_q/m^*$) since

$$\sigma_{OSC}=C(B,T)[\sigma_{xx}(B=0)]\tau_q^2\exp(-\pi/\omega_c\tau_q) , \quad (3)$$

if $\omega_c\tau_q<1$.[8] In Eq. (3), ω_c is the cyclotron frequency, $\sigma_{xx}(B=0)$ the conductivity at zero magnetic field, $C(B,T)$ a weak function of B and temperature, and τ_q the quantum lifetime (Landau-level broadening), which is related to the classical Drude lifetime τ_c by[5]

$$\tau_q/\tau_c=(\pi/4)\tan(\beta/2) , \quad (4)$$

where $\beta=\pi/N_h$ and N_h is the Landau index.[9] For the samples in this work, N_h is about 10, so $\tau_q/\tau_c\approx 0.11$. The ratio τ_q/τ_c is not significantly changed during ir quenching so that the change of τ_c is proportional to that of τ_q.

For a short-gated sample, $\sigma_{xx}(B)$ is directly measured from I_{SD}.[10] For the VDP samples, $\sigma_{xx}(B)$ is calculated from measured values of the resistivities $\rho_{xx}(B)$ and $\rho_{xy}(B)$ by the relation

$$\sigma_{xx}(B)=\rho_{xx}(B)[\rho_{xx}^2(B)+\rho_{xy}^2(B)]^{-1} . \quad (5)$$

Figure 6 shows an example of the dHS oscillations from sample C at 2 K in the following sequence: in dark, after white light, and after IRQ. Shifts in both amplitudes (σ_{OSC}) and periodicity (N_{2D}) are visible. The 100% increase of σ_{OSC} after IRQ is accompanied by a 6% decrease of N_{2D}. This is different from a sample in the dark exposed to white light, which involves macroscopic PPC effects from the GaAs buffer layer; in that case both N_{2D} and σ_{OSC} increase (Fig. 6 where N_{2D} increases by 20%).

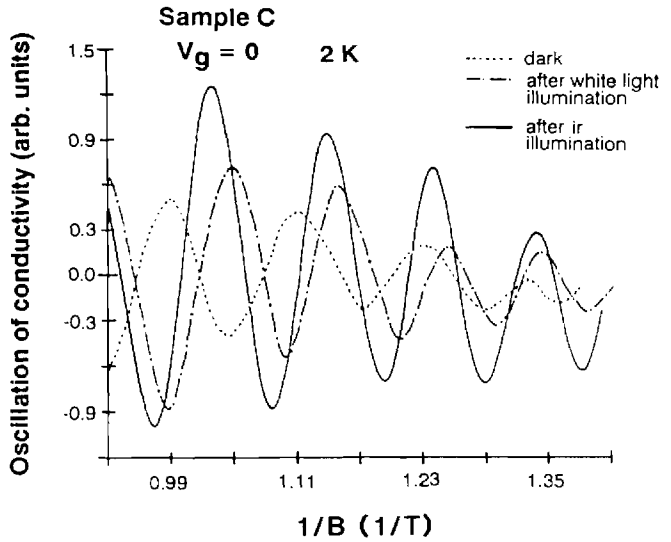


FIG. 6. de Haas–Shubnikov oscillations in the magnetoconductivity from HEMT sample C ($V_g=0$) at 2 K in the illumination sequence of dark, after white light exposure and after ir quenching.

The quantum lifetime is measured by fitting the oscillation amplitude to the leading exponential term in Eq. (3). Typical values of τ_q after ir quenching are

$$\tau_q=2.8\times10^{-13} \text{ sec (sample C)} ,$$

$$\tau_q=3.8\times10^{-13} \text{ sec (sample H)} .$$

dHS measurements complement bulk I_{SD} and N_s measurements since they single out N_{2d} and μ_{2D} at any point during PPC or IRQ, independent of the charge state of the parallel $Al_xGa_{1-x}As$ layer. Suppositions inferred from models or bulk transport can be checked independently.

The relative changes of N_{2D} for samples C and F and ungated samples H and G are summarized in Table II. Samples G and F show 4 and 20% decreases, respectively, in N_{2D} after IRQ (no σ_{osc} data is available because of the low mobility of these samples). Samples H and C ($V_g=0$) show 4.5–6.0% decreases in N_{2D} after IRQ. This is accompanied by an increase of 60–100% in the dHS oscillation envelope amplitude at $B=0.8$ T. Notice that the 4.5% decrease of N_{2D} upon ir illumination from sample H is consistent with the 5% decrease in N_s from the Hall measurement establishing the validity of Eq. (1). The *increase* in N_{2D} from ir illumination for sample C with V_g close to V_t will be discussed in Sec. V.

The data show an increase in N_{2D} from white light and then a decrease upon ir illumination. It could be suggested that white light spreads photoexcited electrons from the gap between the gate and Ohmic contacts to under the gate. To avoid a large potential difference along the channel from this hypothetical charge redistribution (which is not observed), N_1 under the gate would have to decrease proportionately. This is contradicted by transconductance measurements that show V_p shifted to more negative values. The conclusion we reach is rather that free carriers are excited in the $Al_xGa_{1-x}As$ layer under the gate upon white-light illumination and then decrease in that region due to infrared illumination.

We will now show how changes in μ_{2D} can be related to changes in the remote ionized donor population due to PPC and IRQ. It has been determined that remote ionized donors in the $Al_xGa_{1-x}As$ layer provide the dominant scattering mechanism of the 2D EG in $GaAs/Al_xGa_{1-x}As$ heterostructures at low temperatures,[5] so that a decrease in N_i reduces scattering and increases the 2D EG mobility. It has also been shown[11,12] that the classical mobility μ_c of the 2D EG obeys the following relation:

$$\mu_c\approx CN_{2D}^{3/2}d_{eff}^3N_i^{-1} . \quad (6)$$

Here, C is a constant and d_{eff} is the effective distance between the ionized donors and the 2D EG determined by the distribution of ionized donors. The 1.5 power relationship between μ_c and N_{2D} is due to screening effects:

TABLE II. Results from dHS oscillation measurements. The percentage change of the 2D electron number densities (N_{2D}) measured after ir (<0.67 eV) illumination from the value of N_{2D} that remains in the PPC state. The percentage change of the dHS oscillation amplitude σ_{osc} after IRQ at $B=0.8$ T. For samples H and C ($V_g=0$) we present the estimated percentage change of ionized donor density N_i in the $Al_xGa_{1-x}As$ layer following ir quenching of the PPC state, the calculated percentage change of 2D EG classical mobility μ_c including screening effects, the calculated percentage change of the quantum lifetime $(\Delta\tau_q/\tau_q)_c$ from the data of μ_c, N_{2D}, and σ_{osc}. The percentage change of quantum lifetime τ_q obtained by fitting the experimental data directly to the leading exponential term is also presented.

	VDP	HEMT			HEMT		
	H	G	F	F	C	C	C
x	0.23	0.30	0.30	0.30	0.21	0.21	0.21
V_g (V)	ungated	ungated	+0.20	0.0	0.0	−1.95	−2.14
$\Delta N_{2D}/N_{2D}$	−4.5%±1%	−4%±2%	−20%±5%	−20%±5%	−6%±2%	+10%±4%	+50%±5%
$\Delta\sigma_{osc}/\sigma_{osc}$	+60%±10%				+100%±10%		
$\Delta N_i/N_i$	−14%±3%				−13%±4%		
$\Delta\mu_c/\mu_c$	+7%±3%				+4%±3%		
$(\Delta\tau_q/\tau_q)_c$	+9%±3%				+16%±8%		
$\Delta\tau_{rvq}/\tau_q$	+6%±2%				+20%±4%		

a decrease of screening by a reduction in N_{2D} enhances the scattering. There are two limiting cases. (i) Near V_t, the relative change of N_{2D} is much bigger than the relative change in N_i and so μ_c decreases with decreasing N_{2D}. (ii) For incomplete depletion of the $Al_xGa_{1-x}As$ layer both N_{2D} and N_i can change. By the depletion approximation and Poisson's equation $N_i \sim N_{2D}^2$. While white light can contribute a fraction α to the total N_{2D} via electron-hole separation in the GaAs buffer layer, ir light has insufficient energy to induce electron-hole recombination in the GaAs buffer layer. Only $(1-\alpha)N_{2D}$ carriers are depleted from the doped $Al_xGa_{1-x}As$ layer by white light and, from the depletion approximation for a fixed conduction-band discontinuity, a small relative change in N_i can be approximated as

$$\Delta N_i/N_i \approx 2(1-\alpha)^{-1}\Delta N_{2D}/N_{2D} \ . \tag{7}$$

Therefore, for IRQ, the relative decrease in $N_{2D}^{3/2}$ is less than the relative decrease in N_i. Ignoring weak changes in d_{eff}, we observe for incomplete depletion of N_1 that μ_c in Eq. (6) increases when N_{2D} is decreased via IRQ.

The portion α of N_{2D} contributed by electron-hole separation in the GaAs buffer layer may be deduced from the difference between N_{2D} following white light and ir quenching (assuming all photoelectrons originating from traps in the $Al_xGa_{1-x}As$ layer are recaptured after IRQ). For sample H, $\alpha=35\%$, and for sample C ($V_g=0$), $\alpha=10\%$. The resulting estimated values of $\Delta N_i/N_i$ and $\Delta\mu_c/\mu_c$ are shown in Table II.

Two independent methods were used to verify the predicted increase of μ_{2D} in samples H and C ($V_g=0$) both with parallel conductance in the $Al_xGa_{1-x}As$ layer. From Eq. (3), σ_{osc} is an increasing function of both $\sigma_{xx}(B=0)$ and τ_q for fixed magnetic field and temperature. From experimental data showing increases in σ_{osc} after ir quenching and the calculated changes in $\sigma_{xx}(B=0)\approx eN_{2D}\mu_c$, we obtain the changes in τ_q. The results listed as $(\Delta\tau_q/\tau_q)_c$ in Table II indeed show increases for samples H and C ($V_g=0$). The second method directly fits the leading exponential term in Ando's expression to the dHS oscillation data. The results [shown as $(\Delta\tau_q/\tau_q)$ in Table II] also show τ_q increasing after ir illumination. These changes are consistent with the calculated increases of τ_q deduced from σ_{osc}.

In conclusion, dHS oscillations show a decrease in N_{2D} from ir illumination, accompanied by increases in the dHS oscillation amplitude and in τ_q when the $Al_xGa_{1-x}As$ layer is not fully depleted. This suggests ionized traps being filled in the $Al_xGa_{1-x}As$ layer. Together with shifts of V_p during ir illumination, the data indicate that IRQ occurs by the capture of ir-excited free electrons from both the 2D EG and $Al_xGa_{1-x}As$ layer back to ionized donors in the $Al_xGa_{1-x}As$ layer.

IV. CAPTURE CROSS SECTION OF ir-EXCITED ELECTRONS

In the following we use the experimental data to estimate the capture cross sections of ir-excited electrons. It is presumed the ionized donors in the $Al_xGa_{1-x}As$ layer that recapture the carriers are *DX* centers. Under steady infrared illumination, the increase in the density of *occupied DX* centers (N_{DX}) corresponds to a decrease in the free-carrier density, and can be expressed as

$$d(N_{DX})/dt = \sigma(N_{2D0}-N_{DX})(N_{DX0}-N_{DX}) \tag{8}$$

with $\sigma=k(v/l)w$, where k is the trap-capture cross section, v the average z-direction velocity of ir-excited carriers, l is the $Al_xGa_{1-x}As$ layer thickness such that l/v gives the transit time of an ir-excited electron across the $Al_xGa_{1-x}As$ layer, and w is the probability per second of a free carrier being excited by ir illumination of a certain intensity. N_{2D0} is the initial total 2D EG density, such that $(N_{2D0}-N_{DX})$ is the total areal free-carrier density. N_{DX0} is the initial total *unoccupied* areal *DX*-center density at the beginning of an ir quenching interval so that $(N_{DX0}-N_{DX})$ is the unoccupied areal trap density at a given time. A simple single-rate exponential decay would occur when a finite number of carriers were recaptured

by an infinite or very large number of traps, i.e., if $N_{2D0} \ll N_{DX}$. Typically, N_{DX0} is comparable with N_{2D0}, and Eq. (8) is not a simple rate equation, which is consistent with the observed nonexponential decay.

During an infinitesimal time interval δt, the term $(N_{2D0}-N_{DX})$ may be considered as relatively constant in comparison with $(N_{DX0}-N_{DX})$, because $N_{2D0} \gg N_{DX0}$. With this in mind $N_{DX}(t_0+\delta t)$ can be obtained from $N_{DX}(t_0)$ and the initial values of N_{2D0} and N_{DX0} by the following approximation:

$$N_{DX}(t_0+\delta t)=N_{DX}(t_0)\exp\{-\sigma[N_{2D0}-N_{DX}(t_0)]\delta t\} + \sigma N_{DX0}[N_{2D0}-N_{DX}(t_0)]\delta t \ . \quad (9)$$

Equation (9) shows that IRQ occurs as a combination of exponential and linear terms. Both terms have decreasing decay rates as the IRQ progresses because N_{DX} increases during decay. The experimental results behave as predicted and the details of the IRQ can be obtained from Eq. (9).

Assuming that all DX centers are unoccupied at $t=0$ (saturated PPC state), calculated values of $[1-N_{DX}(t)/N_{DX0}]$ for $0<t<100$ sec using Eq. (9) are shown in Fig. 7(a), where $N_{2D0}=10\times10^{11}$ cm^{-2}, for two choices of N_{DX0} and three choices of σ. The decay of $[1-N_{DX}(t)/N_{DX0}]$ is shown in Fig. 7(b), with $N_{DX0}=2\times10^{11}$ cm^{-2} while σ and N_{2D0} were varied. Numerical calculations using Eq. (9) show the normalized decay is sensitive to σ and N_{2D0} but not to changes of the estimated value of N_{DX0}. This is expected for $N_{DX0} \ll N_{2D0}$.

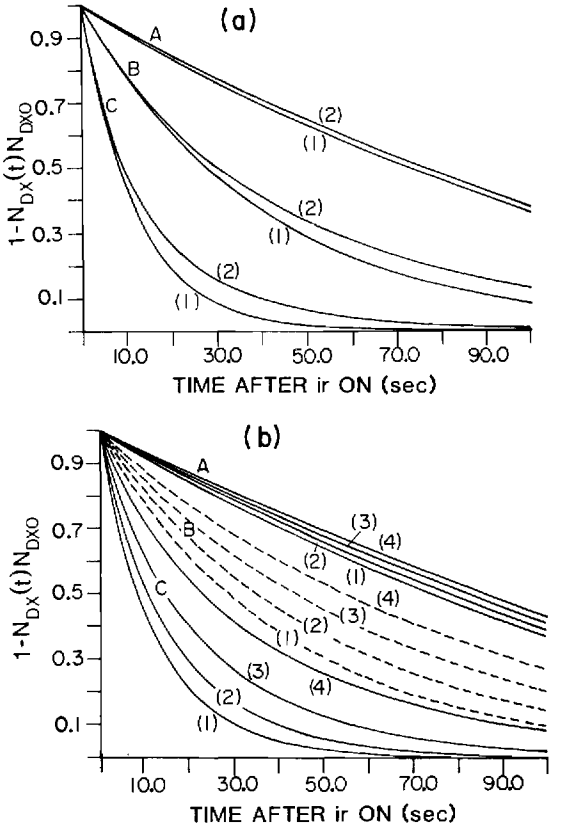


FIG. 7. Theoretical calculation of normalized ir-decay curves (a) shows the weak N_{DX0} dependence [(1) $N_{DX0}=0.5\times10^{11}$ cm^{-2} and (2) $N_{DX0}=5\times10^{11}$ cm^{-2}] but strong σ dependence. σN_{2D0} is fixed at 10×10^{11} cm^{-2}. (b) shows both σ and N_{2D0} are important. N_{DX0} is fixed at 2×10^{11} cm^{-2} and σ and N_{2D0} varied [(1) $N_{2D0}=10^{11}$ cm^{-2}, (2) $N_{2D0}=8\times10^{11}$ cm^{-2}, and (3) $N_{2D0}=6\times10^{11}$ cm^{-2}]. Parameters are chosen to set the values of quenching time close to the experimental range in this work. The values of σ in both (a) and (b) are A, $\sigma=5\times10^{15}$ n^{-2}; B, $\sigma=2.5\times10^{-14}$ cm^{-2}; C, $=8.5\times10^{-14}$ cm^{-2}.

Because of the slow quenching rate in the infinite-time limit, it is assumed that all IRQ-related traps are refilled in the time limit $\sim10^3$ sec, so that N_{DX0} can be determined by the total change of the free-carrier density. All other decay mechanisms are ignored because they have nearly reached their base line values before ir illumination is begun (Figs. 2 and 3).

In order to fit Eq. (9) to the I_{SD} data discussed later, it is necessary to establish the relation between the decay in I_{SD} and the decay of $[1-N_{DX}(t)/N_{DX0}]$, and also determine the N_{2D0} from experimental results. Table III shows results obtained by the methods summarized below.

A. Case i: V_g near threshold (Fig. 4, bottom)

For V_g chosen so as to entirely deplete the free carriers in the $Al_xGa_{1-x}As$ layer, $N_{2D0}-N_{DX}=N_{2D}$. From Eqs. (2) and (6), I_{SD} is considered proportional to $(N_{2D})^{5/2}$ because the relative change in N_i is much smaller than the relative change in N_{2D}. N_{2D0} and N_{DX} are determined by the initial N_{2D} values from dHS oscillations, the I_{SD} data, and the $\frac{5}{2}$ power law relation between I_{SD} and N_{2D}.

B. Case ii: $V_g=0$, parallel conductance dominant (Fig. 4, top)

N_{2D0} and N_{DX} can be estimated for the opposite extreme at $V_g=0$, with parallel conductance in the $Al_xGa_{1-x}As$ layer [$N_1 \gg N_{2D}$, cf. Eq. (2)]. The change in N_{DX} is nearly the same as that in N_1. Since μ_{2D} has been shown to increase when N_{2D} decreases (Table II) during ir quenching, the first term in Eq. (2) for I_{SD} will be essentially constant, and N_1 becomes more important than N_{2D}. The change in I_{SD} during IRQ behaves as a linear function of N_{DX} (since $N_{DX0} \ll N_d$, the change in μ_1 is only weakly dependent on N_{DX} and can be ignored). Before using Eq. (2) to estimate N_1, N_{DX}, and N_{2D0} ($=N_{2D}+N_1$, at $t=0$) from I_{SD} data, the values of μ_{2D}, N_{2D}, and μ_1 are also needed. N_{2D} ($=8\times10^{11}$ cm^{-2}) is directly measured from dHS oscillations. The value of μ_{2D} ($=2.2\times10^4$ cm^2/V sec) is directly measured on sample B ($V_g=0$) by the geometric-magnetotransconductance (GMT) method[3,13] and the value of μ_1 is estimated[4] to be 0.1×10^4 cm^2/V sec. For sample B at 77 K, N_{DX0} at $V_g=0$ is about 2.3×10^{11} cm^{-2}, corresponding to a bulk density of 2.3×10^{16} cm^{-3}, which is much smaller than N_d, and the change in μ_1 can be ignored during ir quenching.

TABLE III. Effective ir-quenching capture cross sections. The estimated value of N_{2D0} (initial total 2D EG number density), or effective N_{2D0} (Ref. 14) and N_{DX0} (initial *unoccupied DX*-center density) of sample B at 77 K (with different gate biases). These were obtained as the sample approached the PPC baseline following a short 1.4-eV light exposure but (see Figs. 2 or 3) before ir ($h\nu < 0.67$ eV) quenching. All densities are in units of 10^{11} cm^{-2}. The fitting values of σN_{2D0} and estimated values of σ ($=kv/lw$) are also presented. The units of σN_{2D0} are sec^{-1}, and the units of σ are 10^{-15} sec^{-1}/cm^2. Even with the variation from the choice of N_{2D0} or effective N_{2D0} an order of magnitude value of σ was determined.

V_g (V)	N_{2D0}	Effective N_{2D0}	N_{DX0}	σN_{2D0}	σ
0	24±4	24	2.4±0.4	0.02±0.005	6–10
−0.15	15±4	24	2.4±0.4	0.02±0.005	6–20
−0.30	7.6±2	24	2.3±0.4	0.03±0.005	12–40
−0.40	6.4±1	24	2.3±0.4	0.03±0.005	12–47
−0.50	2.4±0.4	24	1.5±0.4	0.04±0.01	17–170

C. Case iii: Intermediate situation, $V_p < V_g < 0$ (Fig. 4, middle)

V_g does not change N_{2D} when the $Al_xGa_{1-x}As$ layer is not entirely depleted ($V_g > -0.38$ V for sample B).[3] For $V_g = 0$, −0.4, and −0.5 V (Table III) N_{DX0} has a weak gate-bias dependence until V_g is near V_t and rapid band bending in the $Al_xGa_{1-x}As$ layer narrows the region where IRQ occurs. The agreement in estimated values for N_{DX0} indicates that the assumed relation between I_{SD} and the carriers in different layers versus V_g is a good approximation. The values of N_{DX0} for $V_g = -0.15$ and −0.30 V which are mixed situations (carriers in the $Al_xGa_{1-x}As$ layer and also in the 2D EG) are obtained by linear interpolation. The initial value of N_{2D0} for $V_g = -0.15$ V is the same as for $V_g = 0$, but in when $V_g = -0.30$ V (and $V_g > V_p = -0.28$ V) after IRQ, the $I_{SD} \sim N_{2D}^{5/2}$ power-law relation was used. These results for sample B are summaried in Table III.

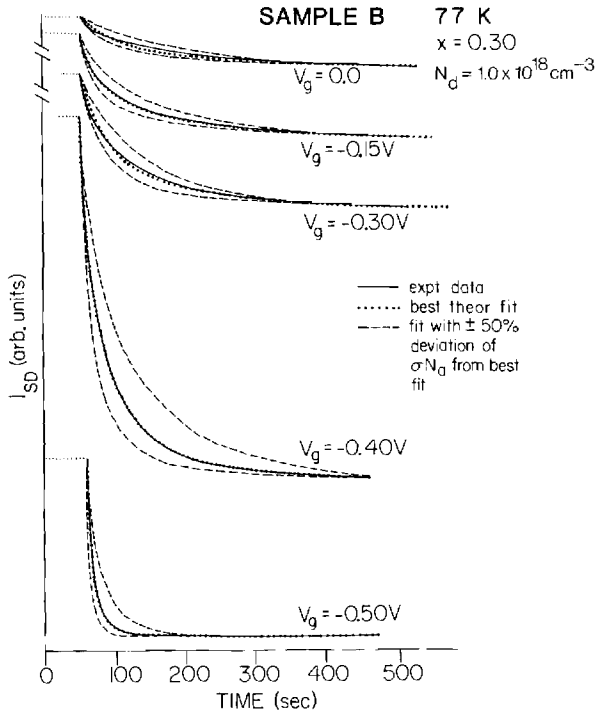


FIG. 8. The theoretical fitting of the ir-quenching data from sample B at 77 K by Eq. (4) using N_{2D0} and N_{DX0} from Table III. The results of σ ($=kv/lw$) are shown in Table III. In addition, Eq. (4) is plotted using values of σN_{2D0} that are ±50% of the best fit values. The I_{SD} data for different gate biases is not shown in same scale, and the I_{SD} zero has been offset for better observation.

Knowing the relation between I_{SD} and N_{DX}, the fit of Eq. (9) to experimental data of sample B at 77 K versus V_g (Fig. 8) gives values of σN_{2D0} and σ that are summarized in Table III. The excellent agreement with experiment for this very nonexponential decay is strong evidence for the validity of the theory of Eq. (9). The decay curves of I_{SD} that are calculated using σN_{2D0} with a 50% deviation from the best-fit values are also presented in Fig. 8. Because the decay is very sensitive to the choice of σN_{2D0} (Fig. 7), variations such as those due to changes in the power law between I_{SD} and are N_{DX} neglected. Although there is a big difference between N_{2D0} and the so-called effective N_{2D0} as V_g approaches V_t it is sufficient to obtain the magnitude of σ, which is on the order of 10^{-14} sec^{-1} cm^2.

Applying this analysis to sample D yields similar values of σ. By fitting Eq. (9) to the ir-quenching curve of N_s from sample D (Fig. 2), and for simplicity, considering that N_s is a linear function of N_{DX}, we obtain $\sigma N_{2D0} = 0.03$ sec^{-1} with $N_{2D0} \approx 1 \times 10^{12}$ cm^{-2} including N_I in the parallel conductance layer. The value of σ for sample D is also on the order of 10^{-14} sec^{-1} cm^2.

D. Absolute ir capture cross section

The ir-light source was a quasiblackbody lamp filtered through a germanium filter with an intensity of ~10 photons sec^{-1} cm^{-2} at the sample. The ir-absorption cross section of free electrons is about 10^{-19} cm^2 in the experiment (see Appendix), so only 1 in 10^7 of the free photoelectrons per second are active in the retrapping process. The dependence on V_g of the length l over which the ir-excited electrons move may be neglected. For example, the distance from the 2D EG to the far e

of the undepleted $Al_xGa_{1-x}As$ layer is 40 nm for $V_g=-0.40$ V and 50 nm for $V_g=0$. This difference makes a negligible contribution to the enhancement of σ. Using $v\approx10^8$ cm/sec (an electron with 0.1–0.5 eV of energy), $l=10^{-5}$ cm and $\sigma[=k_{ir}(v/l)w]\sim10^{-14}$ $sec^{-1}cm^2$, the absolute ir capture cross section, k_{ir} should be on the order of 10^{-20} cm^2. This result for k_{ir} is much smaller than the trap size (about 10^{-15} cm^{-2}) in the $Al_xGa_{1-x}As$ layer, and consistent with the existence of a capture barrier. However, k_{ir} is much larger than the thermal capture cross section (k_T) detected at the same temperature. An estimate of k_T can be reached using $k_T\approx1/(Nv\tau)$. Reference 14 shows that for $x=0.3$ the thermal capture lifetime, $\tau\approx1$ sec at 77 K with a thermal prefactor of 10^{-11} sec, and the *DX*-center capture barrier to be about 150 meV). Using a carrier number density $N\approx10^{18}$ cm^{-3} (same order as doping density), $v\approx10^8$ cm/sec and $\tau=1$ sec, we find $k_T\approx10^{-26}$ cm^2 which is consistent with other reported results.[15] The capture lifetime at a higher temperature corresponds to that of an electron with a larger energy, which is consistent with IRQ because of the much higher effective temperature ($T\approx10^3$ K) of the ir-excited electrons. This capture cross section enhancement by a factor of 10^6 from ir illumination should be considered in reasonable agreement with the results of thermal capture experiments. Our ir-quenching time constants at $T=2$–4 and 30 K show *no significant difference* from the values determined at $T=77$ K for the same sample and gate biases. The temperature independence of the capture cross section is further evidence of the high-energy temperature-independent behavior of the ir-excited electrons.

V. GATE-BIAS DEPENDENCE OF INFRARED QUENCHING

For certain conditions of gate bias we have observed IRQ occurring at energies as low as 0.5 eV and as high as to 1.2 eV (samples B and F). We propose variation in optical threshold is due to internal electric fields (E) which effect the photo threshold of the deep centers. It is possible in gated samples to vary E in the $Al_xGa_{1-x}As$ layer, with E being strongest at the GaAs-cap–$Al_xGa_{1-x}As$ interface (see Fig. 4). Using parameters in Table I and Poisson's equation, the maximum value of E at the interface between the GaAs cap and $Al_xGa_{1-x}As$ layer can be calculated for a given V_g. Samples with a maximum $E<2\times10^5$ V/cm show strong IRQ (Table II). When V_g approaches V_t for sample C (which has a low threshold voltage, $V_t=-2.4$ V after PPC), IRQ weakens until $E\sim3\times10^5$ V/cm ($V_g=-1.6$ V) in the $Al_xGa_{1-x}As$ layer, after which ir light *increases* I_{SD} significantly (Fig. 9). This is observed over the entire temperature range $2<T<77$ K. dHS measurements confirm this is due to an increase of N_{2D} (shown also in Table II).

Figure 10 shows that for the proper choice of V_g (Fig. 4, bottom) ir illumination also *enhances* I_{SD}. The intervals A–D in Fig. 10 show strikingly different rise times to ir illumination depending on net light exposure history for this region of V_g near pinchoff. The data show when the sample has been cooled down in the dark (A), or I_{SD} is near the in-dark value following a long decay interval (D), the rise time is slower than at other times (B,C) in its light exposure history. A depletion calculation (analogous to the differences shown in Fig. 4 from top to bottom), for cases A and D shows that band bending in the doped $Al_xGa_{1-x}As$ layer is much less steep (and E is smaller) than right after white light has been removed. This may explain the slow rise times for these cases. On

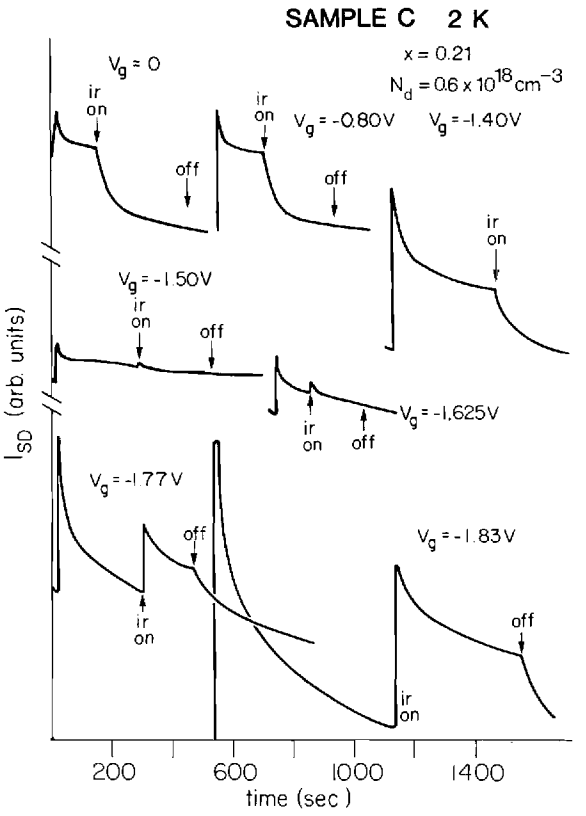


FIG. 9. Gate-bias dependence of ir quenching from sample C with $x=0.21$. When V_g approaches threshold, the ir quenching weakens until the maximum electric field in the $Al_xGa_{1-x}As$ layer is about 3×10^5 V/cm ($V_g=-1.5$ V), after which the ir light actually *increases* the conductivity.

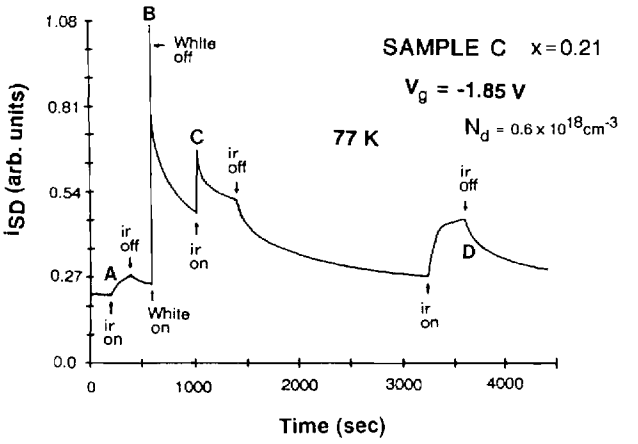


FIG. 10. The ir enhancement with different illumination and decay histories from sample C with $V_g=1.85$V. A–D mark strikingly different rise times to ir illumination depending on net light exposure history for this range of V_g near pinchoff.

the other hand, the slow rise time is not observed when V_g is further from V_t (Fig. 9). An investigation of these results will be left for future work.

In all cases, these ir (or more precisely, the high-energy portion of ir induced *enhancements* of I_{SD} decay with a lifetime of 10^2 sec (for both $T=2$ and 77 K) after the ir light is turned off. This decay is much faster than the typical *DX*-center-related PPC decay (hours to days) at the same temperatures.

A change in the emission rate of trapped electrons due to the Frenkel-Poole effect and phonon-assisted tunneling has been shown to be important in the strong electric field of a heavily doped junction.[16] One such deep trap (the EL2 level in GaAs) with a gate-bias-dependent emission rate has been shown to have *DX*-center-type behavior.[17] For these *DX*-center-like traps (Fig. 11), strong electric fields may also change the energy difference between different configurations. When the microscopic orientation of the *DX* center is favorable, the shift of emission energy from $E_{e,1}$ to $E_{e,2}$ in Fig. 11 (by a change in the applied electric field) must be accompanied by a shift in capture energy from $E_{c,1}$ to $E_{c,2}$, as well as the onset optical threshold from $E_{op,1}$ to $E_{op,2}$. This decrease in the onset optical threshold can explain the change from ir quenching to ir enhancement. The smaller capture energy barrier (or larger capture cross section) under a strong electric field is consistent with the V_g-dependent results of k_{ir} (Table III). But the E-field dependence of the configuration energy of *DX*-center traps is not clear and more detailed experiments should be undertaken in order to understand this in terms of the various models raised for *DX* centers.

When sample C has all its traps in the $Al_xGa_{1-x}As$ layer emptied by cooling down from 300 K at $V_g < V_t$ (at 300 K),[14] the value of V_t at 77 K (-4.0 V) is the same as that at 300 K. In this case ($T=77$ K) there is significant IRQ for $V_g > -3.0$ V, but there is no change in I_{SD} (neither enhancement nor quenching) observed from ir illumination after white light (no PPC effect either) when V_g is set close to V_t because of an absence of trapped electrons. This phenomenon provides additional evidence that under very strong electric fields the emptied traps in the $Al_xGa_{1-x}As$ layer do not capture free electrons.

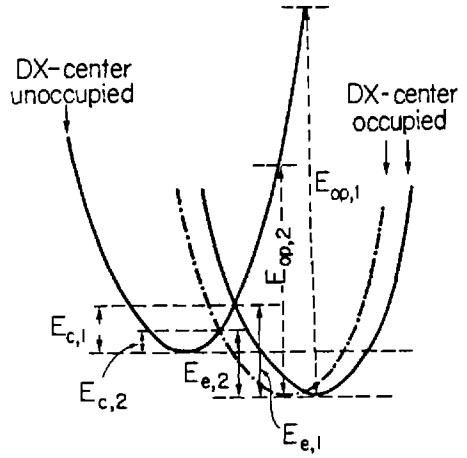


FIG. 11. The proposed electric-field effect on the *DX* centers.

VI. SUMMARY AND FUTURE WORK

This work investigates *DX*-center-related ir quenching via carrier number density and conductance measurements on $GaAs/Al_xGa_{1-x}As$ heterostructures at $2<T<77$ K. The study was carried out on a variety of gated and ungated samples from different sources. We confirm that, after brief white-light exposure to induce the persistent photocurrent state, infrared (ir, $h\nu<0.7$ eV) illumination induces quenching of PPC on ungated samples or gated samples at small gate biases. The commonality of results among these show that the IRQ effect is the same over a range of Al fraction, spacer layer thickness, doping levels, and threshold voltages (see Table I).

Measurements of dHS oscillations show a decrease of the 2D EG number density after ir illumination, accompanied by increases in the dHS oscillation amplitude and the carrier scattering lifetime when the $Al_xGa_{1-x}As$ layer is not fully depleted. This suggests that ir light fills ionized traps in the $Al_xGa_{1-x}As$ layer, which are believed to be the main scattering centers. Together with shifts of the transconductance peak during ir quenching, the data indicate that the quenching is dominated by capture of ir-excited free electrons back to deep centers in the $Al_xGa_{1-x}As$ layer. The ir-quenching times at $2<T<77$ K are *temperature independent*. The ir capture cross section is estimated to be on the order of 10^{-20} cm^{-2}. This result is reasonable considering the energy of ir-excited free electrons and the deep-trap-capture barrier height and decay prefactor.

Experiments on a sample with a higher internal electric field ($>3\times10^5$ V/cm) from V_g show a *strong* increase in the conductivity with the same ir illumination. This E-field dependence of the optical threshold may be due to a change in the trap configuration energy under a strong electric field. More detailed experiments, for example, on the electric-field dependence of the emission barrier from the same traps in the $Al_xGa_{1-x}As$ layer, and the ir photon-energy dependence of the capture cross section of ir quenching, should be undertaken to choose between various models for *DX* centers.

ACKNOWLEDGMENTS

The authors would like to thank R. Gleason and R. Koyama (Tektronix); P. R. Jay and P. Delescluse (Thomson-CSF); W. Schaff (Cornell); and S. C. Palmateer (General Electric) for providing the samples used in this study. This work was supported in part by the Tektronix Foundation and the National Science Foundation through Grant No. DMR85-19728.

APPENDIX

The infrared light in most of this work came from a tungsten lamp passing through a germanium filter with a cutoff energy at 0.67 eV. The tungsten lamp was treated as a quasiblackbody with 10–20 % efficiency and a maximum spectral emittance at 1 μm. About 20% of the to-

tal emitted energy occurs in the 2–5-μm range. For an estimated input power of 10 W to the lamp there should be ~ 0.1 W of infrared (2–5 μm) output. Finally, there was an additional factor-of-10 intensity attenuation from the Ge filter and glass dewar walls.

Theory and experiments have shown[18] that free-carrier absorption coefficient in a semiconductor is proportional to the free-carrier number density, and is inversely proportional to the conductivity mobility. For the near-infrared light (2–5 μm) in this work, the free-electron absorption coefficient is also proportional to the cube of the free-space wavelength when impurity scattering is dominant. From the experimental data on a Te-doped GaAs sample,[19] the absorption coefficient is measured to be about 1 cm^{-1} at 80 K for a free-electron number density $\approx 1\times 10^{18}$ cm^{-3}. This gives a free-carrier absorption cross section on the order of 10^{-18} cm^2. For the MBE-grown samples in this work with free-carrier mobilities larger by a factor of 10, the free-electron absorption cross section is estimated to be on the order of 10^{-19} cm^2.

[1]M. I. Nathan, T. N. Jackson, P. D. Kirchner, E. E. Mendez, G. D. Pettit, and J. M. Woodall, J. Electron. Mater. **12**, 719 (1983); L. X. He, K. P. Martin, and R. J. Higgins, Bull. Am. Phys. Soc. **31**, 656 (1986).

[2]R. J. Nelson, Appl. Phys. Lett. **31**, 351 (1977).

[3]J. P. Harrang, R. J. Higgins. R. K. Goodall, R. H. Wallis, P. R. Jay, and P. Delescluse, J. Appl. Phys. **58**, 4431 (1985).

[4]T. J. Drummond, W. T. Masselink, and H. Morkoç, Proc. IEEE **74**, 777 (1986).

[5]J. P. Harrang, R. J. Higgins, R. K. Goodall, P. R. Jay, M. Laviron, and P. Delescluse, Phys. Rev. B **32**, 8126 (1985).

[6]J. P. Harrang, Ph.D. dissertation, University of Oregon, 1984 (unpublished).

[7]T. J. Drummond, W. Kopp, R. Fischer, H. Morkoç, R. E. Thorne, and A. Y. Cho, J. Appl. Phys. **53**, 1238 (1982).

[8]T. Ando, J. Phys. Soc. Jpn. **37**, 1233 (1974).

[9]D. W. Terwilliger and R. J. Higgins, Phys. Rev. B **7**, 667 (1973).

[10]A. C. Beer, *Galvanomagnetic Effects in Semiconductors* (Academic, New York, 1963), Vol. 71.

[11]P. J. Price, Surf. Sci. **113**, 199 (1982).

[12]K. Lee, M. S. Shur, T. J. Drummond, and H. Morkoç, J. Appl. Phys. **54**, 6432 (1983).

[13]P. R. Jay and R. H. Wallis, IEEE Trans. Electron Devices Lett. **2**, 265 (1981).

[14]P. M. Mooney, N. S. Caswell, P. M. Solomon, and S. L. Wright, in *Microscopic Identity of Electronic Defects in Semiconductors*, MRS Conf. Proc. No. 46, edited by Noble M. Jackson, Stephen G. Bishop, and George B. Watkins (MRS, Pittsburgh, 1985), p. 403.

[15]P. M. Mooney, P. M. Solomon, and T. N. Theis, in *International Symposium GaAs and Related Compounds*, edited by B. de Cremoux, IOP Conf. Proc. No. 74 (IOP, Bristol, 1985), Chap. 7.

[16]G. Vincent, A. Chantre, and D. Bois, J. Appl. Phys. **50**, 5484 (1979).

[17]G. Vincent, D. Bois, and A. Chantre, J. Appl. Phys. **53**, 3643 (1982).

[18]B. A. Smith, *Semiconductors* (Cambridge University, London, 1978), p. 294.

[19]I. Balslev, Phys. Rev. **173**, 762 (1968).

Persistent photocurrent decay mechanisms by capture of photoelectrons in GaAs-$Al_xGa_{1-x}As$ heterostructures

L. X. He, K. P. Martin,* and R. J. Higgins*
Department of Physics, University of Oregon, Eugene, Oregon 97403
(Received 9 February 1987; revised manuscript received 28 May 1987)

We introduce a method for separating multiple-rate decay mechanisms of persistent photoconductivity (PPC) in GaAs/$Al_xGa_{1-x}As$ heterostructures. The derivative of the decay of measured conductivity or carrier number density with respect to the logarithm of time (1 msec–1000 sec) shows a slowly varied base-line structure with some pronounced peaks. The temperature dependence of the varied base line agrees with the temperature dependence of the tunneling barrier. The temperature dependence of the base-line structure and the positions of the pronounced peaks was analyzed to yield capture energies and corresponding lifetime prefactors. The collected evidence demonstrates that the combination of measured capture energies and lifetime prefactors can distinguish the microscopic contributions to PPC decay in the doped $Al_xGa_{1-x}As$ layer of samples from other capture mechanisms. (1) The short-lifetime prefactor (10^{-8}–10^{-10} sec) and associated capture energies are in good agreement with the results from other work on the *DX*-center capture mechanism. (2) This short-lifetime prefactor in a narrow range of temperature compares with prefactors longer by as much as 10 orders of magnitude in adjacent ranges of temperature, which can be identified as tunneling related PPC decay of the two-dimensional electron gas into shallow or deep donors in the $Al_xGa_{1-x}As$ layer. (3) The mechanism with short-lifetime prefactors cannot lie on the GaAs side of the heterojunction because of the observed strong gate-bias dependence and doping-density dependence of the decay magnitude.

I. INTRODUCTION

Persistent photoconductivity (PPC) is known for many semiconductor materials and configurations.[1] Light-enhanced electrical conductivity, carrier concentration, and carrier mobility persist for times ranging from hours to days after the illumination is removed as long as the temperature is below about 100 K, and shorter-term memory is observed even at room temperature. It is important to understand PPC effects in modulation-doped materials and high-electron-mobility transistors (HEMT's) at temperatures below 100 K at which HEMT's display optimal electronic properties. An understanding of PPC effects also contributes to the control of the mechanisms behind current collapse (a change from high to low conductivity following biasing past a certain threshold) and backgating effects (the sensitivity of conductivity in one transistor structure to bias in an adjacent structure) which may have related origins. While in bulk materials or single-layer devices there is some agreement that PPC is related to photoionization of deep levels which have a recombination barrier, a diverse collection of models has been proposed to explain the interplay between *macroscopic* (band-bending) and *microscopic* (deep-level) mechanisms in the layered HEMT structure. This paper is a contribution to separating the dominant mechanisms for PPC-associated decay in HEMT structures from among the two distinct approaches which have been suggested (summarized qualitatively in Fig. 1).

1. Macroscopic-barrier (band-bending) models. The macroscopic-barrier theory[2] involves the spatial separation of photogenerated electrons and holes by built-in electric fields from macroscopic potential barriers due to band bending at surfaces, heterostructure interfaces, or doping profiles. In the case of the $Al_xGa_{1-x}As$/GaAs

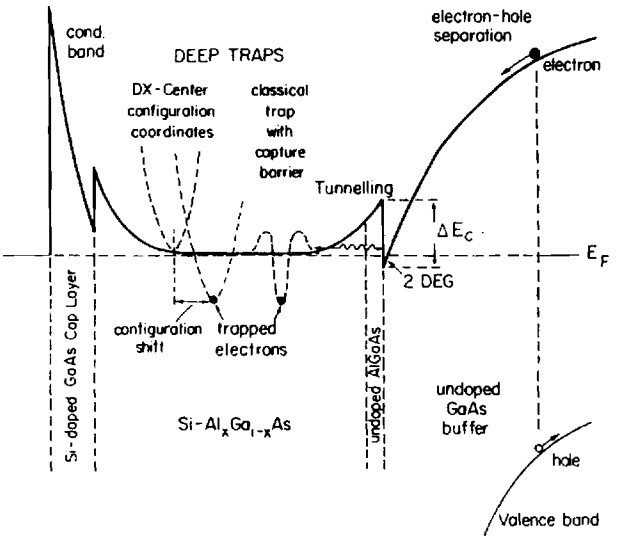


FIG. 1. Major models for persistent photocurrent in HEMT structures, including macroscopic-barrier (band-bending induced electron-hole separation) model and microscopic-barrier (deep-level) models.

heterostructure, this persistence mechanism takes place predominantly in the GaAs side, where the built-in field configuration provides both a mechanism to sweep out photogenerated holes and an electron collection region, the two-dimensional electron gas (2D EG) of the HEMT. Recombination requires that an electron be thermally activated to reach an available hole or hole trap in the GaAs substrate or flat-band region, and requires a time τ:

$$\tau=\tau_r\exp(E_B/kT)\ , \tag{1}$$

where τ_r is the recombination time without the potential barrier, and is determined by other relatively temperature-insensitive decay processes in the recombination sequence, and E_B is the height of the macroscopic barrier. If these effective potential barriers E_B are sufficiently high in comparison to kT, the recombination time τ can become extremely long, since

$$E_B/kT=E_{\text{gap}}(\text{GaAs})/2kT$$
$$=(1.4\text{ eV})/(0.02\text{ eV})=70$$

at 100 K, for example. If traps are present in the substrate or epitaxially grown GaAs buffer region, a distribution of lower values of E_B can be anticipated.

The electron-hole separation mechanism is not considered in the $Al_xGa_{1-x}As$ layer and the cap layer. In this case, the separated holes will move towards either the 2D EG or metal gate (or surface in the ungated case) where some electrons states are always present.

Tunneling[3] also provides a mechanism for electrons to move between the $Al_xGa_{1-x}As$ and GaAs layers despite the macroscopic barrier presented by the conduction band. The lifetime of the tunneling-assisted capture is determined by the tunneling probability and the local capture cross section after tunneling. The local capture cross section of ionized donors is strongly temperature dependent for the case of deep traps, but is only weakly temperature dependent for the case of shallow traps. When kT is much smaller than the height of the tunneling barrier, the Wentzel-Kramers-Brillouin (WKB) method shows the tunneling probability only depends on the shape of the tunneling barrier.

Consider a typical tunneling barrier $U=U(z)$ shown in Fig. 2, where the 2D interface is on the [100] direction and the ionized donors are located in layers parallel to the 2D interface. In the case of $Al_xGa_{1-x}As$, these atomic donor layers are of distance $d=0.283$ nm from each other. After tunneling but before capture, an electron has the same energy as at the 2D interface. For a fixed temperature, the capture after tunneling would be the same for different tunneling distances assuming the local ionized donors have the same capture cross section. Under this assumption, the difference in the tunneling probabilities of an electron reaching different layers will determine the difference of the overall decay lifetime. By the WKB approximation, the probability of an electron reaching the layer $z=z_0$ (Fig. 2) is (the kinetic energy of the electron being ignored)

$$P=\exp\left[\int_0^{z_0}(4\pi/h)(2m^*U)^{1/2}dz\right]\ , \tag{2}$$

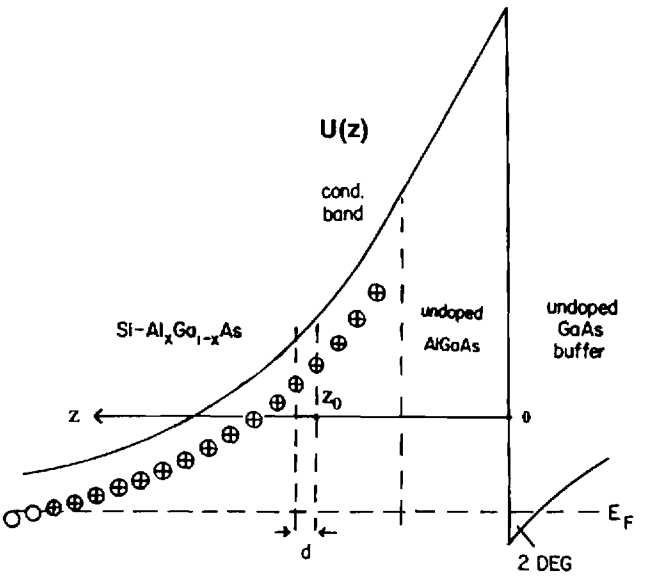


FIG. 2. Typical tunneling barrier for a GaAs/$Al_xGa_{1-x}As$ heterostructure.

where m^* is the effective mass of an electron in the $Al_xGa_{1-x}As$ layer. Because the separation d of atomic planes in the donor layer is much smaller than the typical depletion width (30 nm), and assuming that the overall barrier is nearly the same for an electron reaching the layer at $z=z_0+d$ as for $z=z_0$, then the ratio of the lifetime of an electron captured by the ionized donors in adjacent atomic layers is

$$R=\exp[(4\pi/h)(2m^*U_0)^{1/2}d]\ , \tag{3}$$

where U_0 is the potential barrier height at these adjacent layers. Setting $m^*=0.092m_0$ for an Al fraction $x=0.3$, we have

$$R=\exp(0.879U_0^{1/2})\ , \tag{4}$$

where U_0 is in eV. For example, when U_0 is 0.15 eV, R is 1.4.

When the shape of the tunneling barrier is known (by the depletion approximation), the lifetime for tunneling-assisted capture into each donor layer can therefore be expressed in terms of the lifetime of a 2D EG electron captured by an ionized donor in the first layer from the 2D EG.

Due to the small tunneling probability through the undoped spacer layer between the GaAs and doped $Al_xGa_{1-x}As$ (typically 10^{-4} for a barrier with a 0.3 eV height and 5 nm width) the lifetime of tunneling-assisted capture (even to reach the first layer of ionized donors) is much longer than the decay lifetime by direct capture assuming the same local capture cross section for the ionized donors. The direct-capture lifetime by a shallow donor is of the order of 1 msec at $T=100$ K, but the lifetime for direct capture by a deep donor is of the order of 1000 sec or more at the same temperature. Therefore, for temperatures below 100 K, the observable tunneling-assisted capture through the undoped

$Al_xGa_{1-x}As$ layer within the time window of 1–1000 sec is likely to be shallow-donor related.

2. Microscopic-barrier (deep-level) models. Microscopic-barrier models[4,5] are based on postulated atomic-scale barriers that suppress recombination. The carriers in this case are photoexcited from the impurity center, a deep-donor complex (predominantly identified as the *DX* center in the Si-doped $Al_xGa_{1-x}As$ case) with photon energies smaller than the band gap of the material. These deep donors are thought to be impurity-atom-plus-defect complexes with large lattice relaxations, yielding large Stokes shifts: e.g., $E_{\text{optical}}=0.8$ eV while $E_{\text{thermal}}=0.15$ eV for $Al_xGa_{1-x}As$ with $x=0.35$. A configuration-coordinate model describes this situation. The empty deep level lies above the conduction-band minimum while the occupied deep level lies relatively deep within the gap with a large lattice relaxation required to shift between states. Recent work[6–10] has shown that such *DX* centers are tied to the *L* minimum of the $Al_xGa_{1-x}As$ conduction band. Under this picture, the apparent thermal capture energy must include the energy difference between the Γ and *L* minima.

A formula of the form (1) can also be used to estimate the microscopic barrier overcome during the retrapping of electrons by DX centers. The activation energy in formula (1) in this case is the capture energy E_c, and the prefactor τ_0 is mainly determined by the capture cross section in the high-temperature limit and was shown to be on the order of 10^{-10} sec for $0.27<x<0.35$ from a capture experiment.[11] Recapture dynamics due to deep-level microscopic barriers differ by orders of magnitude from those associated with macroscopic barriers and tunneling, which could also follow a decay curve of the form (1) but with a much longer lifetime prefactor. For example, in the case of shallow-donor capture after tunneling, the lifetime prefactor has been shown[3] to be of the order of 10 sec.

PPC decay kinetics need not be simply exponential. Unfortunately, a variety of mechanisms yield nonexponential decay, so that its observation does not identify a mechanism unambiguously.

(1) Deep-level transient spectroscopy (DLTS) measurements[12] show that both shallow centers and deep centers (*DX*) exist in $Al_xGa_{1-x}As$ layers, with a strong crossover from predominantly shallow donors to predominantly *DX* donors at about $x=0.2$. The *DX* center is found to have a distribution of energies,[13] so that activated capture will not necessarily yield simple exponential decay of PPC.

(2) For the $Al_xGa_{1-x}As$/GaAs heterostructure, Schubert *et al.*[3] showed that PPC decays as a straight line when plotted against the logarithm of time and related this to their tunneling model. They found that the decay curve of a heterostructure at 4 K fits the tunneling model very well.

(3) Queisser and Theodorou[14] also predict a decay that is logarithmic in time due to the spatial distribution of distances between photoelectrons and recombination traps in the GaAs substrate layer.

In view of the variety of mechanisms available, extra information is needed to distinguish among decay mechanisms that contribute to the decay curve. In the work below, we demonstrate how our observed temperature dependence of decay time can be analyzed using a $\ln(t)$-derivative technique to yield a trap spectroscopy whose characteristic energies and times differ so significantly that direct capture (prefactors of 10^{-10} sec) and tunneling mechanisms (with prefactors of 1 sec) are unambiguously identifiable.

II. LOGARITHMIC DERIVATIVE OF MULTIPLE DECAY

The observation of multiple-rate decay behavior led to the use of a more sophisticated signal-processing method[15] in order to separate the temperature dependence of the decay due to each different mechanism.

Assume a general form of multiple-rate decay of the free-electron number density $N(t)$ at a fixed temperature T:

$$N(t)=\sum_i N_{i0}\exp(-t/t_i)\ , \tag{5}$$

where N_{i0} is the initial number density of decay channel i, and with a corresponding individual lifetime t_i.

The key results can be obtained with a more general multiple-rate form with an individual lifetime (t_i):

$$N(t)=\sum_i F(-t/t_i)\ . \tag{6}$$

Define a quantity $K(t)$, or so-called $\ln(t)$ derivative of N, by

$$\begin{aligned}K(t)&=-dN/d[\ln(t)]=-tdN/dt\\&=\sum_i (t/t_i)[dF(-t/t_i)/d(-t/t_i)]\ .\end{aligned} \tag{7}$$

For the moment consider the decay process over a narrow-enough temperature range T to T' so only one decay mechanism dominates, and all t_i can be expressed as (1) with same capture energy. This allows one to write down the following relation between t_i' and t_i:

$$t_i'=t_i\exp[(E/kT')-(E/kT)]\ , \tag{8}$$

where E is the capture energy, and t_i' is the new lifetime at T'.

At temperature T', replacing the lifetimes t_i in (7) by the new lifetimes t_i', we have a new $\ln(t)$ derivative of $N(t)$:

$$K'(t)=\sum_i (t/t_i')[dF(-t/t_i')/d(-t/t_i')]\ . \tag{9}$$

The time axis can be changed from t to a new time axis t' (i.e., make a change of time units) by

$$t=t'\exp[(E/kT')-(E/kT)]\ . \tag{10}$$

Notice that the individual lifetime t_i' is not altered by the change of the time axis. It is obvious from the above that $t/t_i'=t'/t_i$. Replacing t/t_i' in (9) by t'/t_i, we find that in the new time axis at temperature T'

$$K'(t)=K'(t')=\sum_i (t'/t_i)[dF(-t'/t_i)/d(-t'/t_i)]\ . \tag{11}$$

The right-hand side of (11) is the same form as (7) except that t is replaced by t'. A plot of (7) with $\ln(t)$ will be identical with a plot of (11) with $\ln(t')$. Because $\ln(t)=\ln(t')+[(E/kT')-(E/kT)]$, the plot of (9) with $\ln(t)$ must be simply displaced by $[(E/kT')-(E/kT)]$ from the plot of (7) with the same $\ln(t)$. Therefore, the capture energy E and the decay-time prefactor can be determined from the displacement measured by experiment.

For the simplest case of (5) with only a single decay lifetime t_0, such as direct capture by ionized donors in the doped supply layer,

$$K(t)=(t/t_0)N_{00}\exp(-t/t_0)\ . \tag{12}$$

The function is strongly peaked at precisely t_0, as shown in Fig. 3(a), with a peak value of N_{00}/e. We can measure the decay magnitude N_{00} from the peak value of $K(t)$. The half-width of the single decay peak in the $K(t)$-$\ln(t)$ plot is about one decade in time.

For the case of multiple-rate decay with the general form of (5), each decay channel is associated with a similar peak in $K(t)$ as its characteristic decay time. This forms the basis for a multiple-rate decay spectroscopy.

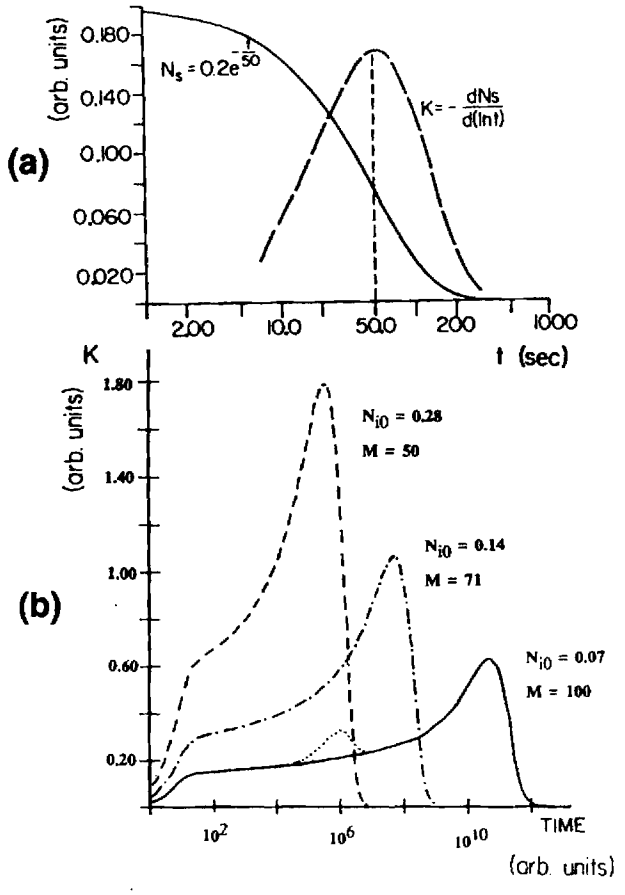


FIG. 3. Multiple-rate trap spectroscopy. (a) Any single exponential decay vs $\ln(t)$; the derivative (dashed line) displays a peak at the characteristic decay time. (b) A calculated result of $K(t)$ vs $\ln(t)$ of tunneling-assisted decay with different refilled ionized-donor densities N_{i0} and total number M of depleted donor-atom layers. The dotted curve at the bottom shows how a single isolated trap would superimpose.

From the definition of $K(t)$, we have

$$K(t)=\sum_i (t/t_i)N_{i0}\exp(-t/t_i)\ , \tag{13}$$

with peaks at $t=t_i^*$, where t_i^* can be slightly different from t_i. But when the assumption $t_i^*=t_i\exp[(E/kT')-(E/kT)]$ holds for a suitable temperature interval (N_{i0} is nearly constant in this temperature interval), from the general proof above, we can see that the peaks at t_i^* with a $\ln(t)$ plot shift from temperature T to a new temperature T' by $[(E/kT')-(E/kT)]$ in $\ln(t)$.

A linearlike plot of $N(t)$ versus $\ln(t)$ is obtained from expression (13) for a multiple-rate decay (5) with a certain form of the distribution of N_{i0}. As an example we consider the case of tunneling-assisted capture with a step-function tunneling barrier. From formula (2), the lifetime τ of an electron reaching a donor-atom layer is proportional to the exponential of the distance (z_0) from the 2D interface to this atomic donor layer. If N_{i0} has a continuous uniform distribution in the z direction in the doped layer, we have $dN=N_0\,d[\ln(\tau)]$ for fixed N_0. For an arbitrary time t within a given time window, when the longest lifetime (capture by the last donor-atom layer) and the shortest lifetime (capture by the first donor-atom layer) is far outside of the range of the time window, we have

$$K(t)=\int_0^\infty (t/\tau)(N_0/\tau)\exp(-t/\tau)d\tau=N_0\ . \tag{14}$$

$K(t)$ becomes a constant, and the plot of $N(t)$ with $\ln(t)$ will be a straight line. This conclusion also holds even for the more general form shown in (11).

For a well-separated multiple-rate decay process, $K(t)$ as determined from experimental decay data will show a sequence of peaks, with each peak corresponding to an individual decay channel in the multiple-rate process. Due to the broad half-width of a single decay channel in a $K(t)$-$\ln(t)$ plot, these peaks can be distinguished experimentally only when the separation is large enough. This condition is not generally satisfied for tunneling-assisted capture, because the difference in capture time between adjacent donor-atom layers is too small.

To investigate the behavior of tunneling-assisted decay with the realistic tunneling barrier, let us consider the simplest tunneling case shown in Fig. 2. Under the depletion approximation, the potential barrier height of the ith donor-atom layer can be expressed as

$$U_i=U_0(1-i/M)^2\ , \tag{15}$$

where M is the total number of the depleted donor atom layers and U_0 is the barrier height at the first layer. Using formula (3), all values of τ_i can be expressed in terms of τ_1, the lifetime of an electron captured by the first donor-atom layer, by the following iterative relation:

$$\tau_{i+1}=\tau_i\exp(0.879U_i^{1/2})\ . \tag{16}$$

Assuming a certain form of the distribution (N_{i0}) of ionized donors refilled by the tunneling electrons, the value of $K(t)$ can be calculated by formula (9).

A numerical simulation is shown in Fig. 3(b) using an arbitrary time unit that is determined by the lifetime of the tunneling-assisted capture into the first donor layer, and can be deduced from experimental data. In the solid curve, we assume 100 layers of donors with the same single-layer ionized-donor number density $N_{i0}=0.07$ (arbitrary units). The value of U_0 is chosen to be 0.3 eV, and the depletion width is the thickness of 100 donor-atom layers. The lifetime for the first layer is set to be 10 (arbitrary units). From the solid curve, we can see that $K(t)$ slowly increases over 10 orders of magnitude in time after initial capture by the first several layers begins until capture by the last few layers. Actually, the tunneling-assisted capture is a superposition of the capture contribution by each layer. The total result is the envelope of the superposition. The increasing slope up to the cutoff time has a physical explanation. When each layer makes the same contribution, the higher barrier at the first few layers in the depletion region nearest the 2D EG will have a bigger difference in the lifetime from adjacent layers. In this case, the peaks will be separated more from each other in a $\ln(t)$ plot, and have smaller overlap, giving smaller net $K(t)$ value. At larger time (further distance from the 2D interface) where the barrier becomes smaller, the small separation between peaks will show increased superposition, giving a higher $K(t)$ value. In this calculation, we ignore the possible nonuniform distribution of the single-layer ionized-donor number density. Otherwise, the real $K(t)$-$\ln(t)$ plot of tunneling-assisted decay may show additional fluctuations. But, with or without these fluctuations, the overall horizontal shift of the plot at the same values of $K(t)$ from different temperatures can be used to determine the activation energy and lifetime prefactor of capture by the local ionized donor. However, in fact, we will see that the temperature dependence of the ionization of donors at higher temperatures (typically above 100 K), and consequently different depletion lengths, will cause an extra shift of the tunneling-assisted decay curve. This is especially important when we interpret results for tunneling into the last few layers. In this case, although the tunneling may only be associated with shallow ionized donors, the decreasing tail of the $K(t)$-$\ln(t)$ plot [Fig. 3(b)] can have an extra shift towards a shorter time because of the decrease of the total number of tunneling layers. The dashed curves in Fig. 3(b) are the plots for the cases with larger $N_{i0}=0.14$ and 0.28 (same arbitrary units) and consequently shorter depletion length with a total of 71 and 50 layers of depleted donors, respectively. This is calculated by assuming that N_{i0} is proportional to the ionized-donor number density and the depletion width is determined by the depletion approximation. The lifetime for the tunneling-assisted capture into the first donor layer is set to be the same for the case of a shallow ionized donor. The plot shows clearly that the value of $K(t)$ for the capture into the first few donor layers is proportional to N_{i0}, and the time where the decreasing tail of the curve begins is strongly dependent on the depletion width. This strong shift is not from the change of the capture lifetime by the local ionized donors. The calculation shows that the amplitude of the central portion of the curve is about 3–10 times as large as the refilled ionized-donor number density of a single donor-atom layer. Considering that the fraction of refilled ionized donors out of the total ionized donors is of the order of a few percent, the amplitude of $K(t)$ for tunneling-assisted capture is of the order of the ionized-donor number density of a single donor-atom layer.

When the lifetime for direct capture into one type of ionized donor and the lifetime of tunneling-assisted capture into possibly another kind of donor is comparable within the same time window, the results of the mixture will depend on the amplitude of each kind of decay. If there are no, or very few, free carriers in the doped layer after illumination, we may not be able to distinguish any single decay peak in the plot of $K(t)$-$\ln(t)$, and tunneling-assisted capture will be the dominant decay mechanism. When their amplitudes are comparable, the slowly increasing base line will be augmented by a single peak [Fig. 3(b), dotted line]. At different temperatures, the shifts of the base line and the peak will be different if they correspond to different types of ionized donors. When the free-carrier density in the doped layer is much larger than the single-layer ionized-donor density, direct-capture decay can dominate, and the slowly increasing base line becomes less significant.

III. EXPERIMENTS AND DISCUSSION

Most of the molecular-beam-epitaxy (MBE) heterostructure samples in this work were composed of the following layer sequence: Si-doped GaAs cap layer, Si-doped $Al_xGa_{1-x}As$ layer, thin undoped $Al_xGa_{1-x}As$ spacer layer, thick undoped GaAs buffer layer, and semi-insulating GaAs substrate. A two-dimensional free-electron gas is formed at the interface between the undoped $Al_xGa_{1-x}As$ and GaAs buffer layers, separated by the spacer layer from the nearest donors in $Al_xGa_{1-x}As$ layer.

Samples from four different sources were used. Summarized in Table I are their principal parameters: aluminum fraction x, cap-layer thickness L_c, doping density $N_{d,c}$ of the cap layer, $Al_xGa_{1-x}As$-layer thickness L_a, $Al_xGa_{1-x}As$-layer doping density $N_{d,a}$, undoped spacer thickness L_s, gate voltage threshold (V_t) at 77 K in the dark, the net ionized-donor density N_i in the depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, and the percentage increase of N_i after identical brief white-light illumination, sufficient to saturate the PPC-induced change. Samples $A-C$ were gated samples. They are classic high-electron-mobility transistors. Samples A and B are on the same chip, and had $x=0.3$. Sample C had $x=0.21$. All metal gates are directly on top of the cap layer, which are not very heavily doped to ensure good gate characteristics. The samples with thick and heavily doped $Al_xGa_{1-x}As$ layers have free carriers in the $Al_xGa_{1-x}As$ layer when the gate bias is set higher than that required to deplete the doped supply layer (like that shown in Fig. 1). In this case, direct capture by an ionized donor will occur in the $Al_xGa_{1-x}As$ layer, and tunneling takes place through the undoped layer. Sam-

TABLE I. Sample label, source, aluminum ratio x, cap-layer thickness L_c, doping density $N_{d,c}$ of the cap layer, $Al_xGa_{1-x}As$-layer thickness L_a, $Al_xGa_{1-x}As$-layer doping density $N_{d,a}$, undoped spacer thickness L_s, gate voltage threshold V_t at 77 K in the dark, the net ionized-donor density N_i in the depleted $Al_xGa_{1-x}As$ layer at 77 K in the dark, and the percentage increase of N_i after identical brief white-light illumination sufficient to saturate the PPC change. Layer thicknesses are in units of nm and doping and carrier densities are in units of 10^{17} cm^{-3}.

Sample	Source	x	L_c (nm)	$N_{d,c}$	L_a (nm)	$N_{d,a}$	L_s (nm)	V_t (V)	N_i	$\Delta N_i/N_i$
A, B	Tek.[a]	0.3	50	2	100	10	5	−0.5	1.9	15%
C	TCSF[b]	0.21	20	6	120	6	8	−2.0	2.9	11%
D, DD	Cornell[c]	0.3	20	10	30	10	10	ungated	VDP	
G	Tek.[a]	0.3	50	2	100	10	5	ungated		
L	Gain[d]	0.3	30	20	50	15	4	ungated	VDP	

[a]Tek.: Tektronix Inc., Read Gleason and Richard Koyama.
[b]TCSF: Thomson-CSF (Paris), Paul Jay.
[c]Cornell: Cornell University, W. Schaff.
[d]Gain: Gain Electronics Corporation, T. Hurl.

ple *D*, an ungated van der Pauw (VDP) square, had $x=0.3$ in the Si-doped $Al_xGa_{1-x}As$ layer but $x=1.0$ in the undoped spacer layer, which makes for an extra-high tunneling barrier (1.0 eV). There is a 10 nm undoped $Al_xGa_{1-x}As$ layer ($x=0.3$) between the cap layer and doped $Al_xGa_{1-x}As$ layer. Sample *DD* has similar layer parameters and sample *L* (an ungated VDP) was chosen as a heavily doped sample in order that there be an extra large number of free carriers in the doped layer.

Gaps of 2–5 μm separate the gate from source or drain in our HEMT samples, and are sufficiently wide enough to prevent the gate voltage from influencing material under source and drain contact regions for small source-drain signals.[16] The source-drain voltage (V_{SD}) is kept below 1 meV to avoid hot-electron effects, and is much smaller than the smallest activation energy (>5 meV) involved.

All gated samples had a short enough gate length (1–2 μm) so that the diffused illumination could reach beneath the gate. This is proven by observed shifts in transconductance curves. In the charge-control picture,[17] the transconductance is low in the region where the gate controls only low-mobility carriers in the $Al_xGa_{1-x}As$ layer, and is much higher in the region where the gate controls the 2D EG number density. The transconductance peak occurs after the $Al_xGa_{1-x}As$ carrier density is depleted and the 2D EG is being modulated, and the pinchoff point occurs when the total carrier density at the 2D interface has been depleted. Shifts of the transconductance peak and the pinchoff threshold during the light illumination therefore give additional information about the changes of carrier densities in the different layers. Data from all gated samples at 77 K show large shifts (≥ 0.3 V) in the transconductance curves from their dark values following white light illumination, which relax back during the decay process towards their original form on very long time scales ($t>10^5$ sec). These measurements (which uniformly show a shift towards increased pinchoff-voltage magnitude) permit us to conclude unambiguously that light reached under the gate to increase the free-carrier density in the $Al_xGa_{1-x}As$ layer as well as the 2D EG.

For gated HEMT samples, the source-drain current I_{SD} was measured with V_{SD} fixed. For ungated van der Pauw samples, the carrier densities N_s were measured via the Hall effect. Both methods involve the number densities in the 2D and $Al_xGa_{1-x}As$ layers (N_{2D} and N_1) in a similar way.[18] When the mobility of 2D electrons (μ_{2D}) is much larger than the mobility of electrons in an $Al_xGa_{1-x}As$ layer (μ_1) such that $\mu_1 N_1 \ll \mu_{2D} N_{2D}$ (we ignore the difference between Hall mobility and conductance mobility because only the ratio μ_1/μ_{2D} is important in our case),

$$N_s = N_{2D} + 2N_1\mu_1/\mu_{2D} \text{ (Hall measurement)}, \quad (17)$$

$$I_{SD} = -eGV_{SD}\mu_{2D}(N_{2D} + N_1\mu_1/\mu_{2D}) \text{ (conductivity measurement)}, \quad (18)$$

where G is a geometrical factor for short-gate samples. In our cases, $\mu_1 \approx 10^3$cm^2/V sec while $\mu_{2D} \approx 10^5$ cm^2/V sec and N_1 never exceeds N_{2D} significantly, such that the condition for Eq. (17), $\mu_1 N_1 \ll \mu_{2D} N_{2D}$, holds. In both measurements of N_s and I_{SD}, the N_{2D} term contributes more than the N_1 term.

Since PPC and its associated decay produces a change in conductivity, not a net charging, a change in the number of free carriers is accompanied by a change in fixed charge. In the simplest decay case, when V_g is near V_t (and the $Al_xGa_{1-x}As$ layer is totally depleted), N_{2D} is small compared to the ionized donor density in the $Al_xGa_{1-x}As$ layer so that tunneling-assisted decay in N_{2D} does not strongly change N_i. This means μ_{2D} decreases from reduced screening with the net result being the decay of I_{SD}. When V_g is chosen so that the $Al_xGa_{1-x}As$ layer is not totally depleted, the observed decay in I_{SD} occurs partially from a decrease of low-mobility carriers via direct recapture to deep levels. Simultaneously N_{2D} self consistently decreases in order to satisfy the depletion approximation, as well as decaying by tunneling-assisted recombination. The decay in I_{SD} results from this combined decrease of N_1 and N_{2D} (although there may be weak increases in their mobilities[7,19]). Nevertheless, the overall relaxation behavior of

I_{SD} of N_s is mostly dependent on the capture energy and lifetime prefactors of the decay mechanisms and not strongly determined by details of the dependence of the mobilities on the charge state of the sample.

Given the persistence range of seconds to hours at the temperatures of interest here, most data were recorded between 0 and 1000 sec so that a decay-rate "window" of 1–500 sec is measured in our experiments. This includes what is sometimes referred to as the "transient" photocurrent regime, but we do not adopt this terminology since the selection of time window merely sets the temperature range for a given photoconductance process to manifest itself, as in deep-level transient spectroscopy.

The decay is nonexponential, as illustrated in Figs. 4(a) and 4(b) from a VDP dc measurement on sample *DD* at 77 K. The time window is 10^{-3}–10^{5} sec. Typically the source-drain current or VDP sheet density decays after brief white-light illumination and shows a fast initial decay, followed by a decay rate that decreases continuously as the decay approaches a base line above the dark value (the long-term PPC). The decay nearly behaves as a straight line when plotted as a function of $\log_{10}(t)$ over nearly 7 orders of magnitude in time. For convenience, the $K(t)$ value calculated in this experiment is with respect to $\log_{10}(t)$. In this way, a single-rate decay with an amplitude of N_{00} will correspond to an experimental result of $K(t)$ with a peak value of 0.85 $\times N_{00}$. The logarithmic time derivative of N_s is also plotted as a dashed curve in Fig. 4(b). The curve of $K(t)$ shows a single decay peak seen at $t = 10$ msec, while

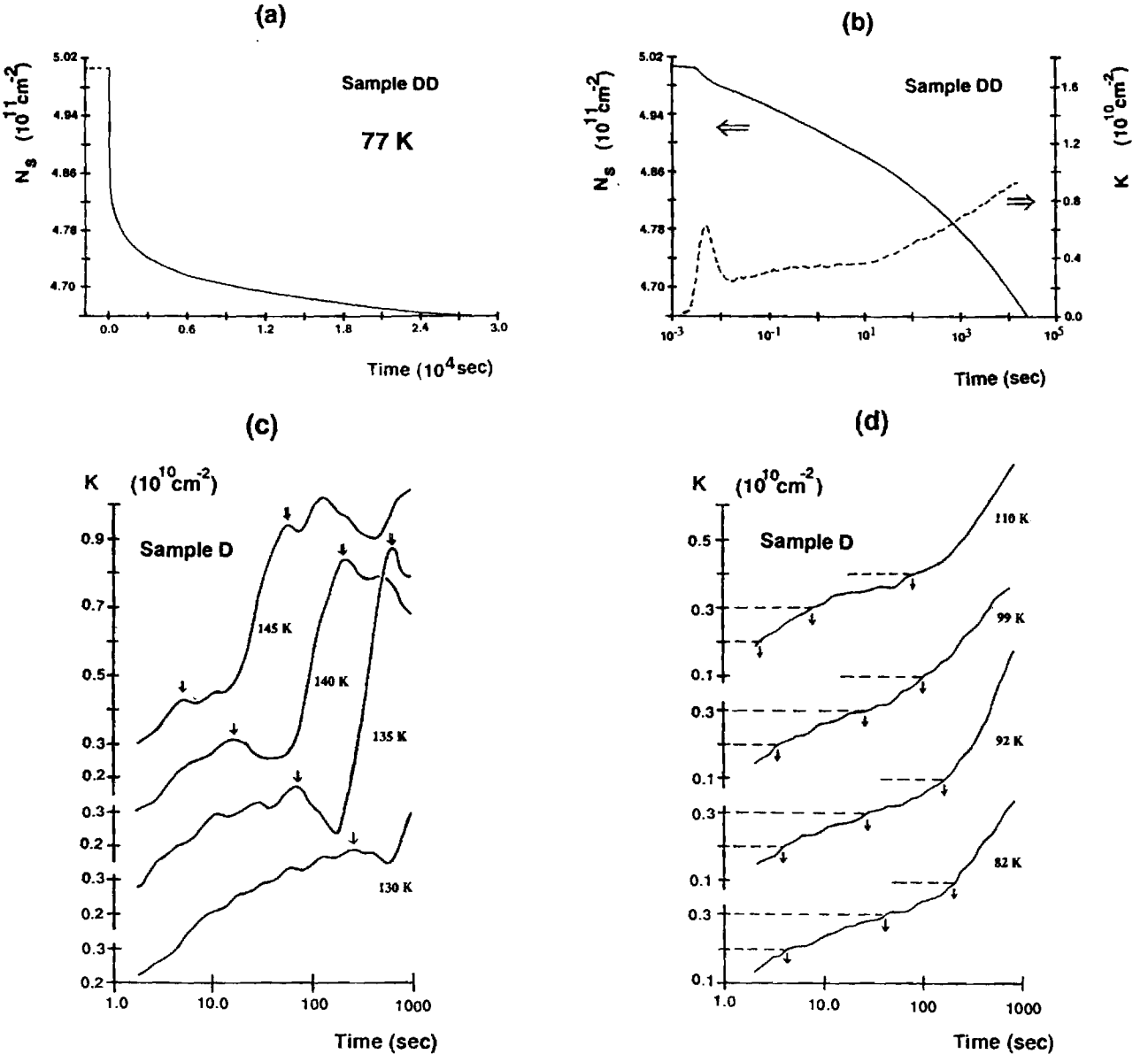


FIG. 4. (a) Nonexponential PPC decay from sample *DD* at 77 K over the 10^{-3}–10^{5}-sec rate window. It is plotted as a function of linear time. (b) The same decay plotted with logarithmic time. The value of $K(t)$ is also plotted as a dashed curve. (c) and (d) are examples of the experimental $K(t)$ data from an ungated van der Pauw square (sample *D*).

most of the curve is a slowly increasing base line over nearly 6 orders of magnitude in time. This behavior is consistent with the theoretical prediction of Fig. 3(b).

In repeated measurements following illumination sufficient to saturate I_{SD} or N_s, the $K(t)$-$\log_{10}(t)$ plot of the decay curve shows no significant shift of these features at the same temperature. Raising the temperature speeds up each decay, but as long as the temperature change is small enough, individual peaks or the structure of curves can be tracked. The shifts of the decay structures with changing T gives the characteristic activation energies of the individual decay features. Figure 4(c) shows the shifts of decay groups with temperature from the measurements of N_s on sample D (ungated). The broad individual decay peaks can easily be tracked on top of a slowly increasing base line in the 130–145-K temperature range. Notice that the peak values of $K(t)$ are nearly the same when they shift from different temperatures. The peak value of $K(t)$ is 1×10^{10} cm^{-2}, which corresponds [from formula (12) plus a factor of $\ln 10=2.30$] to 6×10^{10} cm^{-2} of directly captured electrons. From the temperature dependence of the peak shifts, we can deduce capture energies and lifetime prefactors. The capture energy of these peaks is measured to be 370 meV with a lifetime prefactor of 10^{-10} sec. By contrast, in Fig. 4(d) in the temperature range of 82–110 K, there is no obvious track of peak shifts, but a slowly increasing base-line structure becomes clear, whose overall horizontal shift at the same values of $K(t)$ can be tracked at different temperatures. The same value of $K(t)$ is marked on the decay curves taken at different temperatures. A similar method was used over the range of 145–180 K.

The track of the single decay peaks or the track of the slowly increasing base-line structure [see discussion following formula (11)] from sample D is shown in Fig. 5(a) and summarized in Table II. Figure 5(a) displays several distinct decay mechanisms at different temperature regimes, marked as (ii)–(iv). Over the temperature ranges of 80–125 K (iv), sample D shows a retrapping energy of 20–40 meV, with a prefactor value equal to about 1–100 sec indicating that tunneling into shallow donors is involved. Actually, the tunneling-assisted capture must have the same capture energy as that corresponding to direct capture by the same kind of ionized donor. However, the lifetime prefactor is much longer due to the small tunneling probability. The dominant activation energy increases to about 140 meV with a 10^{-4}-sec lifetime prefactor at higher temperatures $T=145$–180 K (ii) as deeper traps are involved after tunneling and as the macroscopic barrier width decreases due to a thermally induced increase of free carriers in the $Al_xGa_{1-x}As$ layer at higher temperatures.[20,21] This is consistent with the measurements (in the dark, or with illumination) on sample D and sample DD [to be discussed later in Fig. 7(b)] which show that the sheet number density is increasing rapidly when T is greater than 120 K but is nearly constant when T is below 110 K. In addition to tunneling-related retrapping back to deep and shallow donor centers, thermally activated behavior was detected over a narrow intermediate temperature range, as seen in the steep portions [marked (iii)] of Fig. 5(a). Sample D displays a retrapping energy of 370 meV with a 10^{-10}-sec prefactor over a narrow temperature range at 130–145 K [region (iii)], in sharp contrast with the factor-of-3 lower capture energy and 6-orders-of-magnitude longer prefactor at adjacent temperatures (ii) and (iv). Although sample D has an extra-high barrier in the spacer layer, the intermediate values of activation energy yet long (1–100 sec) lifetime prefactor of regions (iv) can be explained by tunneling of persistent photoelectrons between the GaAs cap layer and the adjacent $Al_xGa_{1-x}As$ layer ($x=0.30$), since the conduction band

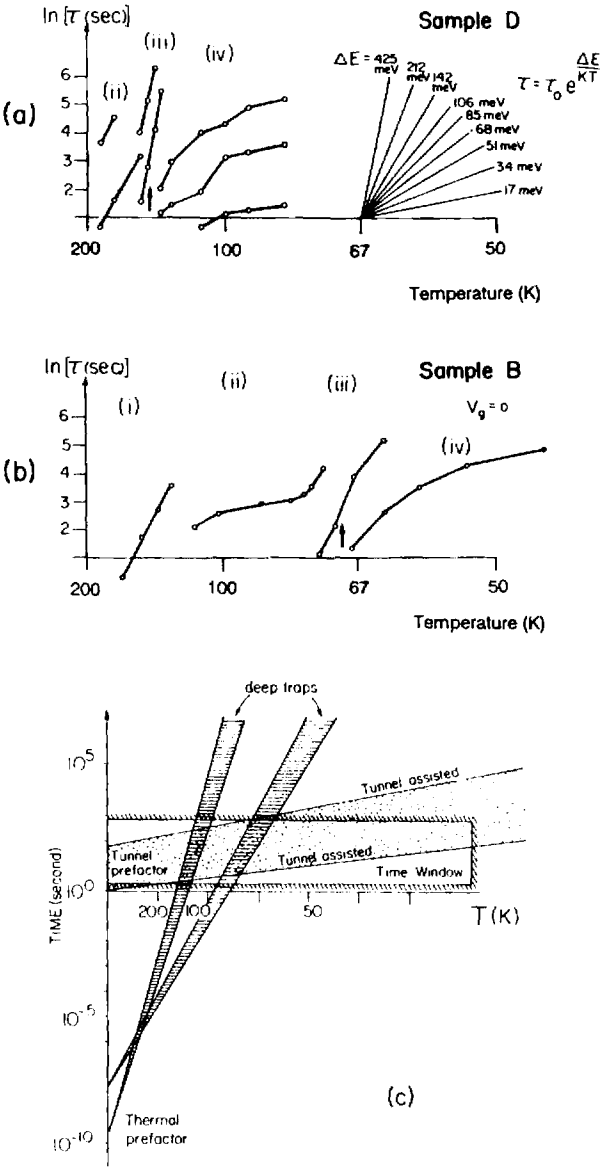


FIG. 5. Temperature dependence of peaks associated with groups of traps from multiple-rate trap spectroscopy. (a) Sample D; (b) sample B ($V_g=0$); (c) sketch of the dominant features of the above results. Data extracted from different samples in different temperature regions labeled in this figure are shown in Table II.

TABLE II. Data extracted from Fig. 5, including the labels of different sample and different temperature regions, the temperature range of the labeled region, the capture energy E_c, deviation of capture energies ΔE_c, and lifetime prefactors τ_0.

Sample	Label in figure	Temperature range (K)	E_c (meV)	ΔE_c (meV)	τ_0 (sec)
B ($x=0.3$)	(i)	120–160	150	20	10^{-4}
B	(ii)	80–120	30	20	5
B	(iii)	65–75	150	20	10^{-8}
B	(iv)	45–60	20	10	5
D ($x=0.3$)	(ii)	145–180	140	20	10^{-4}
D	(iii)	130–145	370	40	10^{-10}
D	(iv)	80–125	30	10	1–100

of the doped GaAs cap layer in this sample is closer to E_F than in the sketch of Fig. 1, and can hold photoexcited free carriers. Since DX centers in very heavily doped GaAs have been detected,[22,23] direct capture by a DX center can occur in a heavily doped cap layer. The results from ungated VDP sample L with a much wider and heavier doped GaAs cap layer and doped $Al_xGa_{1-x}As$ layer show that along with a slowly increasing base line with $K(t)\approx 2\times 10^{10}$ cm^{-2}, a large single decay peak due to direct recapture with $K(t)$ of the order of 4×10^{11} cm^{-2}, which is 40 times larger than the case of sample D, was observed and the reduced capture energy from the peak shifts is 350 meV in the temperature range 140–170 K.

An example from the conductivity measurements on sample B ($V_g=0$) shown in Fig. 5(b) also displays several distinct decay mechanisms in different temperature regimes, marked as (i)–(iv) and summarized in Table II. Over the temperature ranges of 45–65 K (iv) and 80–120 K (ii), sample B shows a retrapping energy of 15–50 meV, with a prefactor value equal to about 1 sec indicating that tunneling into shallow donors is involved. The dominant activation energy increases to about 150 meV at higher temperatures (i) as deeper traps are involved following tunneling and as the macroscopic barrier width decreases due to a thermally induced increase of free carries in the $Al_xGa_{1-x}As$ layer at higher temperatures. In addition to tunneling-related retrapping back to deep and shallow donor centers, thermally activated behavior was detected over a narrow intermediate temperature range, as seen in the steep portions [marked (iii)] of Fig. 5(b). Sample B displays a retrapping energy of 150 meV with a 10^{-8}-sec prefactor over a narrow temperature range at 65–75 K [region (iii)] in sharp contrast with the factor-of-8 lower capture energy and 8-orders-of-magnitude longer prefactor at adjacent temperatures (ii) and (iv). An ungated sample G from the same wafer shows a 170 meV capture energy at $T=80$ K, a slightly higher temperature than observed for gated samples. Sample C ($V_g=-0.8$ V) with $x=0.21$ also clearly displays this sharp kink signature of the high-activation-energy (200 meV) thermally activated process over a narrow temperature range of 65–75 K. The capture energy is 220 meV from sample C with $V_g=0$ over the same narrow temperature range. But for the same sample with V_g close to pinchoff ($V_g=-1.5V$), direct capture by an ionized donor was not observed.

All the dominant features from our samples can be summarized in Fig. 5(c). A high-activation-energy (150 or 370 meV) short-prefactor (10^{-8}–10^{-10} sec) process crosses through the experimental rate window over a narrow temperature range, intersecting a low-activation-energy, long-prefactor process which is visible in the data of Figs. 5(a) and 5(b) as lines of the small slope on either side of the steep-slope portion.

The key conclusion that qualitatively different decay channels may be separated using the $\ln(t)$-derivative technique can be verified in an alternative decay experiment shown in Fig. 6. A continuous temperature sweep

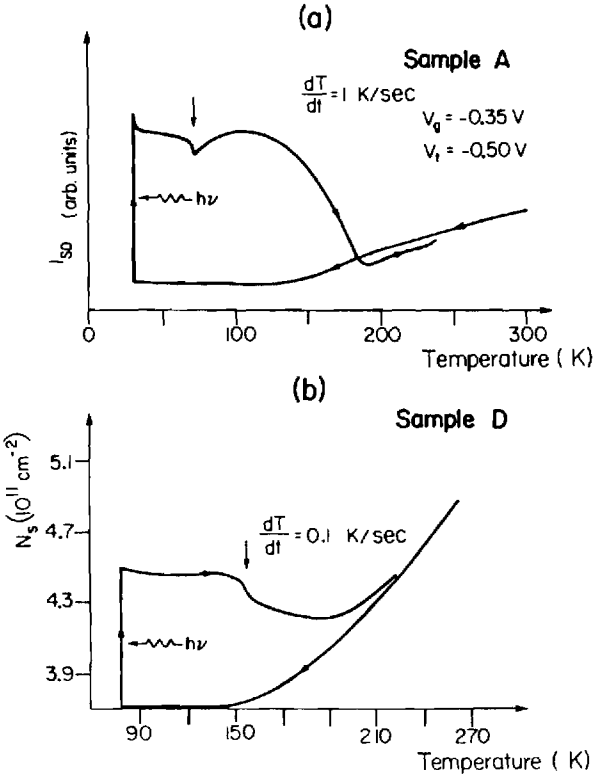


FIG. 6. Temperature-sweep method of observing elevated PPC decay at special temperatures (arrow) in (a) gated transistor structures and (b) Hall patterns.

at a constant rate of about 1 K/sec performed on sample A after illumination is shown in Fig. 6(a). I_{SD} shows a faster decay in the temperature interval near 70 K, in agreement with the kink portion observed in the data in Fig. 5(b) at a comparable temperature. The temperature at which the extra dip occurs of course shifts with sweep rate as in DLTS [shifting rate window in Fig. 5(c)] and is measured to be weakly gate-voltage dependent. A similar result is obtained for sample D in a temperature-sweep measurement of N_s [Fig. 6(b)] when the sample is warmed from 77 K at 0.1 K/sec after a brief white-light illumination. A sudden dip in number density is observed (see arrow) when T passes through 150 K, which agrees with the steep-slope portion of Fig. 5(a).

In this work, we have obtained three different apparent DX-center capture energies with different Al fractions: 370 meV for $x=0$, 220 meV for $x=0.21$, and 170 meV for $x=0.30$. When the energy differences between the L and Γ minima at different Al fractions are included,[6] the net capture barrier height in the local DX-center configuration coordinate model is 90 ± 30 meV for all samples we measured. This is in reasonable agreement with DLTS results for the emission barrier height (200–400 meV) and the fixed energy value (150 meV independent of x) by which the DX center is located below the L minimum of the conduction band.[7,13]

An example of a more complete picture of the tunneling-assisted multiple-rate decay is from dc VDP measurements [data shown in Fig. 7(a)] on sample DD (from the same wafer as sample D). The time window is set from 1 msec to 500 sec. To detect the decay in these short times, a light emitting diode (LED) is used for illumination and the instruments are set to have a 0–10 kHZ response. In the temperature range of 86–178 K, every curve shows a single decay peak (with a few msec lifetime) and is followed by a slowly increasing base line. The temperature dependence of the single short decay lifetime displays the behavior of shallow donors as above. The decay curve that follows can be fit to the numerical tunneling calculation of Fig. 3(b). Sample DD has an extra-high barrier between the $Al_xGa_{1-x}As$ and GaAs layers due to the aluminum ratio $x=1$ in the 10-nm spacer layer, which prevents tunneling through this barrier to be observed in our decay-rate window. Only the tunneling-assisted decay between the cap layer and the doped AlGaAs layer is detected. For simplicity, the measured change of the sheet number density is proportional to the change of the free carriers in the parallel conductance layer. From formula (17), the apparent value $K(t)$ from the measured sheet number density will be smaller than the real value due to capture in the parallel conductance layer. Nevertheless, the basic feature of tunneling-assisted decay can be clearly seen in the experimental results with the slowly increasing value of $K(t)$ and the decreasing tails when the depletion width becomes shorter at higher temperatures.

The dashed curves in Fig. 7(a) are the calculated results using the following parameters. The potential barrier height U_0 at the first ionized-donor layer is calculated from the conduction-band discontinuity (0.23 eV for $x=0.3$) at the 2D interface and the potential drop through the 10-nm undoped $Al_xGa_{1-x}As$ layer (Fig. 2). The potential drop from the 2D interface is readily calculated using the depletion approximation[24] resulting in

$$N_i e/2\epsilon=\Delta E_c(M^2d^2+2LMd)^{-1}, \tag{19}$$

$$U_0=N_i e/2\epsilon M^2d^2, \tag{20}$$

where ΔE_c is the conduction-band discontinuity at the 2D interface, ϵ is the dielectric constant ($\epsilon=13$ for our

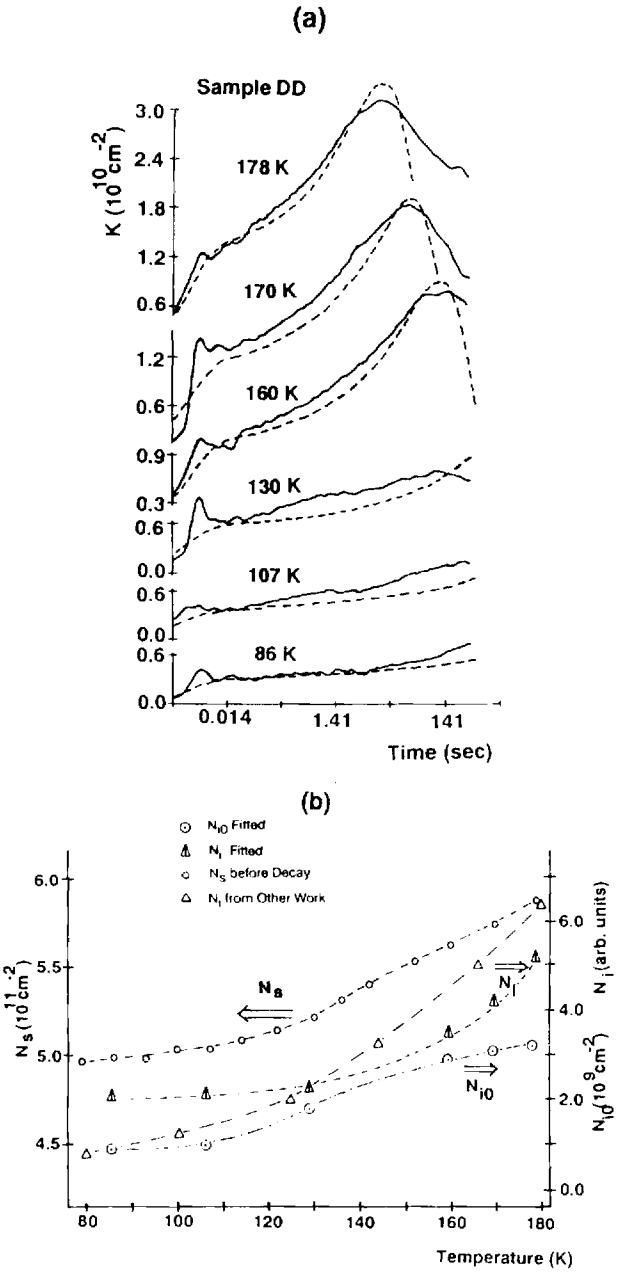


FIG. 7. (a) Experimental results (solid curves) of the tunneling-assisted capture from ungated VDP sample DD. Dashed curves are the calculated results from a realistic tunneling barrier. (b) The parameters N_{i0} and N_i used in the calculation of (a). The experimental data of the sheet number density N_s before each decay, and the temperature dependence of the bulk ionized-donor density N_i from other work is also presented.

TABLE III. The fitted values of N_{i0} (refilled electron density of a single donor-atom layer) and the fitted values of M (the total number of ionized donor-atom layers) from the data in Fig. 7(a). The ionized-donor densities N_i calculated from the value of M are also presented.

T (K)	N_{i0} (10^9 cm^{-2})	M	N_i (10^{17} cm^{-3})
86	0.9	105	2.27
107	1.0	104	2.30
130	1.9	102	2.38
160	2.9	78	3.61
170	3.1	70	4.25
178	3.3	61	5.21

case), L is the thickness of the undoped spacer layer, and M is the total number of ionized-donor-atom layers such that Md is the depletion width. From (19) and (20), we can calculate the value of U_0 and N_i by choosing the value of M. The lifetime of the tunneling-assisted capture into the first layer is selected to be 10 msec (the tunneling barrier in the undoped layer and the lifetime of a shallow donor are not very temperature sensitive). Using formulas (13), (15), and (16) for a fixed conduction-band discontinuity, the overall variation of $K(t)$ with time is determined by the ionized-donor density N_i (consequently the depletion width), but the amplitude of $K(t)$ is determined by the values of N_{i0}. The fitted values of N_{i0} and the corresponding values of M and N_i found from the fit at different temperatures are shown in Table III. The theoretical calculation fits the data fairly well, and the parameters are consistent with the doping density of 1×10^{18} cm^{-3}, which gives a maximum value of $N_i=1\times10^{18}$ cm^{-3} and a corresponding maximum $N_{i0}=2.8\times10^{10}$ cm^{-2}.

This variation in the fitting parameters N_{i0} and N_i requires a depletion width which increases with increasing temperature, which is consistent with the following independent check. The temperature dependence of the parameters N_{i0} and N_i used to fit data in Fig. 7(a) are shown in Fig. 7(b). The measured sheet number density N_s before each decay is also plotted for comparison. The value of N_s is generally an increasing function of N_i in the doped supply layer with increasing temperature, and Fig. 7(b) shows the same behavior for N_s and fitting parameters N_{i0} and N_i. In Fig. 7(b), the temperature dependence of the bulk $Al_xGa_{1-x}As$ ionized-donor density with an Al fraction of $x=0.34$ and doping density of 1.7×10^{17} cm^{-3} is also presented,[25] which also shows a similar temperature dependence.

IV. CONCLUSION

We have introduced a method for separating multiple-rate decay mechanisms in persistent photoconductivity (PPC) decay of GaAs/$Al_xGa_{1-x}As$ heterostructure. The derivative of the decay of measured conductivity, or carrier number density with respect to the logarithm of time shows a slowly varied base-line structure with some pronounced peaks. The temperature dependence of the varied base line agrees with the temperature dependence of the tunneling barrier. The temperature dependence of the base-line structure and the positions of the pronounced peaks were analyzed to yield two distinctly different classes of decay mechanisms: a high-activation-energy, short-prefactor time mechanism associated with the DX center in doped AlGaAs, and a low-activation-energy, long-prefactor mechanism associated with tunneling-assisted recombination of 2DEG electrons across the macroscopic barrier between GaAs and $Al_xGa_{1-x}As$.

The collected evidence from this work demonstrates that the combination of measured capture energies and lifetime prefactors can distinguish the microscopic mechanisms (direct capture) for PPC decay in the doped layer from other capture mechanisms. (1) The short-lifetime prefactor (10^{-8}–10^{-10} sec) and associated capture energies are in good agreement with results from other work on the DX-center capture mechanism.[11,19] These short-lifetime prefactors compare with prefactors longer by as much as 10 orders of magnitude at adjacent temperatures (Fig. 5), which can be identified as tunneling-related photoconductance decay of 2D electrons into shallow or deep donors in the AlGaAs layer. (2) The gate bias does not change the electric field within the GaAs buffer region. So the mechanism with a short-lifetime prefactor cannot lie on the GaAs side of the heterojunction because of the observed strong gate-bias dependence and the doping-density dependence of the decay magnitude.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation, Division of Materials Science (NSF-DMR) Grants No. 81-19550, No. 81-13456, and No. 85-19728, and by the Tektronix Foundation. It is a pleasure to thank K. R. Gleason (Tektronix), P. R. Jay and P. Delescluse (Laboratoire Central de Recherches, Thomson-CSF), W. Schaff (Cornell), and T. Hurl (Gain) for supplying useful samples and J. D. Cohen, M. N. Wybourne, and A. V. Gelatos (University of Oregon) for helpful discussions.

*Present address: School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

[1]M. I. Nathan, Solid-State Electron. **29**, 167 (1986).
[2]D. E. Theodorou and H. J. Queisser, Appl. Phys. **23**, 121 (1980).
[3]E. F. Schubert, A. Fischer, and K. Ploog, Phys. Rev. B **31**, 7937 (1985).
[4]D. V. Lang and R. A. Logan, Phys. Rev. Lett. **39**, 635 (1977).
[5]D. V. Lang and R. A. Logan, Phys. Rev. B **19**, 1015 (1979).
[6]A. K. Saxena, Solid-State Electron. **25**, 127 (1982).

[7]N. Chand, T. Henderson, J. Klem, W. T. Masselink, R. Fischer, and H. Morkoç, Phys. Rev. B **30**, 4481 (1984).
[8]M. Mizuta, M. Tachikawa, H. Kukimoto, and S. Minomura, Jpn. J. Appl. Phys. **24**, L143 (1985).
[9]P. M. Mooney, E. Calleja, S. L. Wright, and M. Heiblum, in *Proceedings of the 14th International Conference on Defects in Semiconductors*, edited by H. J. von Bardeleben (Trans Tech, Switzerland, 1987), Vols. 10–12, p. 417.
[10]E. Calleja, P. M. Mooney, S. L. Wright, and M. Heiblum, Appl. Phys. Lett. **49**, 657 (1986).
[11]P. M. Mooney, N. S. Caswell, P. M. Solomon, and S. L. Wright, Mater. Res. Soc. Symp. Proc. **46**, 403 (1985).
[12]M. O. Watanabe, K. Morizuka, M. Mashita, Y. Ashizawa, and Y. Zohta, Jpn. J. Appl. Phys. **23**, 103 (1984).
[13]P. M. Mooney, R. Fischer, and H. Morkoç, J. Appl. Phys. **57**, 1928 (1985).
[14]H. J. Queisser and D. E. Theodorou, Phys. Rev. B **33**, 4027 (1986).
[15]A. D. N. Swingler, IEEE Trans. Biomed. Eng. **BME-24**, 408 (1977).
[16]R. Fischer, T. J. Drummond, J. Klem, T. Henderson, D. Perrachione, and H. Morkoç, IEEE Trans. Electron Devices **ED-31**, 1028 (1984).
[17]J. P. Harrang, R. J. Higgins, R. K. Goodall, R. H. Wallis, P. R. Jay, and P. Delescluse, J. Appl. Phys. **58** (1985).
[18]J. P. Harrang, Ph.D. thesis, University of Orgeon, 1984.
[19]R. J. Nelson, Appl. Phys. Lett. **31**, 351 (1977).
[20]T. J. Drummond, W. Kopp, R. Fischer, H. Morkoç, R. E. Thorne, and A. Y. Cho, J. Appl. Phys. **53**, 1238 (1982).
[21]E. F. Schubert and K. Ploog, Phys. Rev. B **30**, 7021 (1984).
[22]P. M. Mooney, Bull. Am. Phys. Soc. **32**, 504 (1987).
[23]M. I. Nathan, M. Heiblum, T. N. Morgan, L. Eaves, D. K. Maude, J. C. Portal, R. B. Beall, and J. J. Harris, Bull. Am. Phys. Soc. **32**, 553 (1987).
[24]T. J. Drummond, W. T. Masselink, and H. Morkoç, Proc. IEEE **74**, 777 (1986).
[25]M. O. Watanabe and H. Maeda, Jpn. J. Appl. Phys. **23**, L734 (1984).